University of California
Santa Barbara

Frozen Gaussian Approximation for Elastic Waves, Seismic Inversion and Deep Learning

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Mathematics

by

James Charles Hateley IV

Committee in charge:

Professor Xu Yang, Chair
Professor Hector D. Ceniceros
Professor Carlos Garcia-Cervera

June 2019

The Dissertation of James Charles Hateley IV is approved.

Professor Hector D. Ceniceros

Professor Carlos Garcia-Cervera

Professor Xu Yang, Committee Chair

May 2019

Frozen Gaussian Approximation for Elastic Waves, Seismic Inversion and Deep Learning

by

James Charles Hateley IV

To my wife Meagan for putting up with my antics and my daughter Julia for being Julia.

Acknowledgements

Thanks to Jay Roberts and Kyle Mylonakis for their assistance with neural networks and with the development of the API GeoSeg.

Thanks to Lihui Chai for his guidance and assistance with Fortran and MPI.

Thanks to Xu Yang for his guidance and many useful discussions.

Support is acknowledged from the Center for Scientific Computing from the CNSI, MRL: an NSF MRSEC (DMR-1720256) and NSF CNS-1725797.

This research was also partially supported by the NSF grants DMS-1418596 and DMS-1818592.

Curriculum Vitæ

James Charles Hateley IV

Education

2019	Ph.D. in Mathematics (expected), University of California, Santa Barbara.
2010	M.S. in Mathematics, University of California, Irvine.
2008	B.S. in Mathematics, University of California, Irvine

Publications

1	J. C. Hateley and X. Yang. Convergence of the Frozen Gaussian Approximation for High Frequency Elastic Waves. (In preparation).
2	J. C. Hateley, J. Roberts, K. Mylonakis and X. Yang. Deep Learning Seismic Substructure Detection using the Frozen Gaussian Approximation. Geophys. J. Int. (Submitted). Preprint. http://arxiv.org/abs/1810.06610, 2019
3	J. C. Hateley, L. Chai, P. Tong, X. Yang. Frozen Gaussian approximation for 3-D Elastic Wave Equation and Seismic Tomography. Geophys. J. Int. Volume 216, Issue 2, February 2019, Pages 1394–1412, https://doi.org/10.1093/gji/ggy498
4	J. C. Hateley, H. Wei and L. Chen. Fast Methods for Computing Centroidal Voronoi Tessellations. Journal of Scientific Computing, 2014, (DOI) 10.1007/s10915-014-9894-1.

Abstract

Frozen Gaussian Approximation for Elastic Waves, Seismic Inversion and Deep Learning

by

James Charles Hateley IV

The frozen Gaussian approximation (FGA) is an efficient solver for high frequency wave propagation. This work is to generalize the FGA to solve the 3-D elastic wave equation and use it as the forward modeling tool for seismic tomography with high-frequency initial datum. The evolution equation is derived by weak asymptotic analysis in conjunction with projecting onto an orthonormal frame; this is numerically verified and analytically proven to have same asymptotic error as the eigenfunction decomposition. Examples from seismology are given by forward modeling, solving an inverse problem and generating data sets to train neural networks.

Contents

Curriculum Vitae vi

Abstract vii

List of Figures x

List of Tables xv

List of Symbols 1

1 Introduction 2
- 1.1 Permissions and Attributions . . . 2
- 1.2 Governing Equation . . . 2
- 1.3 WKBJ Approximation . . . 5
- 1.4 Introduction to the FGA . . . 8

2 Frozen Gaussian Approximation 13
- 2.1 Preliminaries . . . 13
- 2.2 Decomposition and initial condition . . . 21
- 2.3 Derivation of the Evolution Equation . . . 24
- 2.4 Proof of Convergence . . . 32

3 Frozen Gaussian Approximation and Seismology 53
- 3.1 Remarks on Seismology and the FGA . . . 53
- 3.2 Interface and Boundary Conditions . . . 55
- 3.3 Greens Function from a point source . . . 63
- 3.4 Seismic Inversion . . . 66

4 Numerical Experiments 71
- 4.1 Techniques for Implementation . . . 71
- 4.2 3D Models . . . 73
- 4.3 Seismic Tomography . . . 83

5 Deep Leaning with the FGA **90**
5.1 Neural Networks and seismology 91
5.2 Numerical Experiments 97

6 Concluding remarks **114**
6.1 Future Research 115

Bibliography **117**

List of Figures

3.1 Cartoon illustration of an incident Gaussian wave packet for P-wave hitting the interface at $z = z_0$, and then reflected and transmitted as Gaussian wave packets for P- and SV-waves. Here the $G_{\mathrm{p,s}}^{\mathrm{in,re,tr}}$ stands for the Gaussian wave packet for the incident, reflected and transmitted P- and SV-waves, respectively. We denote $\theta_i, \theta_r, \theta_t$ to be the incident, reflection and transmission angles of P-waves, and ϕ_r, ϕ_t to be the reflection and transmission angles of SV-waves, respectively. 56

4.1 Illustration of the main three steps of the FGA algorithm: Step 1, decompose the initial wavefields into Gaussian wave packets, and calculate the corresponding weight function $\psi_{\mathrm{p,s}}^k$ defined in eq. (2.37); Step 2, Time propagation, i.e., solve numerically eqs. (2.14), (2.66), (2.34), (2.35) and (2.36) for the dynamics of Gassian center, propagation vector and prefactor amplitude. Transmission and reflection conditions are needed at the presence of interfaces as given in Section 3.2; Step 3, Wave reconstruction, i.e., to compute eq. (2.29) for the elastic wavefield at time T. 72

4.2 Modulus of the elastic wavefield computed by the FGA method in homogeneous media, with the analytical solution given by eq. (3.48). The subfigures from left to right show the slices of $\|\boldsymbol{u}\|_2$ at $y = 64$ km for $t = 0, 6.93, 13.86$ s. The frequency of the source time function is $f = 1.4702$ Hz. 76

4.3 Modulus of the component-wise elastic wavefield computed by the FGA method in homogeneous media, with the analytical solution given by eq. (3.48). The subfigures show the components of $\boldsymbol{u} = (u_1, u_2, u_3)^{\mathrm{T}}$ for $t = 6.93$ s and the source frequency $f = 1.4702$ Hz. The second and third columns show the computed P- and S-wavefields by the FGA method, respectively. 77

4.4 Comparison of accuracy for FGA and SPECFEM3D to the analytical solution (3.48) at $t = 6.93$ s with the source frequency $f = 1.4702$ Hz. The top subfigures are the modulus of wavefields given by FGA, SPECFEM3D and analytical solution. The bottom subfigures are the modulus differences between the FGA and analytical solutions (bottom left) and between the SPECFEM and analytical solutions (bottom right). This shows that FGA and SPECFEM3D produce comparable accuracy in this case. . . . 78
4.5 Dependence of one-step computational time on frequency for both FGA and SPECFEM3D in homogeneous media. The horizontal axis is the frequency f (Hz) and the vertical axis is the one-step computational time (s) of the solvers. The triangle line stands for the FGA simulations and the square line stands for the SPECFEM3D simulations. Due to the limitation of memory, SPEFEM3D can not run for $f \geq 5.88084$ Hz, and the dashed square line is obtained by extrapolation. 79
4.6 Dependence of time on the number of processors for the source frequency $f = 1.4702$ Hz in homogeneous media. The ideal speed-up ratio is 2, while one can see that the speed-up ratio for FGA is approximately 1.9393, which is slightly smaller than 2 indicating an almost perfectly parallel efficiency. On the other hand, as a comparison, SPECFEM3D is used with 128 elements in each spatial direction to achieve a comparable accuracy to FGA. The speed-up ratio for the SPECFEM3D solver is around 1.4828, which is smaller that those of FGA. This is because SPECFEM3D solves eq. (2.2) on a parallel computer with N processors by partitioning the whole domain into N slabs with each processor solving the equation in each slab. Therefore, for each time step, each processor needs to communicate with its neighbors to get necessary boundary information, which decreases the speed-up ratio. 79
4.7 The Marmousi Model (a) P-wave velocity profile for an xz-plane. Stations are in red spanning horizontally at 130 m intervals starting at $(4, 1.008, 0.1)$km. Source is located in black at $(5.9, 1.008, 0.5)$ km. The velocities are in km/s (b) Slice of the Marmousi Model at $y = 1.008$ km with $T_0 = 0.125s$, $\sigma = 0.0442$ s, $f = 40.7436$ Hz 80
4.8 Modulus of P and S wavefields at various time slices. Top row: $t = 0.5$ s (a) Modulus of wavefield. (b) Modulus of P-wavefield. (c) Modulus of S-wavefield: Bottom row: (d) Modulus of wavefield at $t = 1$ s. (e) Modulus of wavefield at $t = 1.5$ s 81
4.9 Seismograms for $T_0 = 0.125s$, $\sigma = 0.0442s$, $f = 40.7436$Hz, receivers located at 130m intervals. 81
4.10 (a): The layered P-wave velocity follows the data in the IASP91 model [51, 47]. (b): Cartoon plot of the 410-km discontinuity, and distributions of source and stations. 82

4.11 The seismic signals of P-waves received at stations of depth 480 km, 420 km and 360 km, respectively, simulated by the finite difference (FD) and FGA methods, with the source locating at the depth of 600 km. FGA shows a good agreement with the reference signals computed by the FD method for the P-waves. 82
4.12 (a): The three-layered crosswell P-wave velocity model with homogeneity in horizontal y-direction. The background velocities from top to bottom are $C_1 = 1800$ m/s, $C_2 = 2000$ m/s, $C_3 = 2200$ m/s, and the interfaces locate at $z_0 = 0$ m, $z_1 = 100$ m, $z_2 = 200$ m. The low-velocity region has a Gaussian shape centered at $x_c = 75$ m, $z_c = 150$ m, with standard deviation equal to 1/30 m. 16 stars indicate the locations of seismic sources, and 32 dots indicate the locations of seismic receivers. (b): we use the relative velocity $\Delta c/c = (c_{\mathrm{p}} - c_0)/c_0$ to indicate the largest magnitude of the low-velocity perturbation from the background velocity and the area of low velocity region. 84
4.13 The signals of direct P, pP, S2P, pS, direct S, and S2pS received at the top station with the source as the top ninth source in Fig. 4.12. Here 2 means the second interface locates at $z_2 = 200$ m, e.g. S2P means S-wave gets reflected at the interface at $z_2 = 200$ m and arrives at the station as P-wave. 86
4.14 Seismic tomography kernels computed by FGA for the initial velocity model (4.7), with the thick dashed lines as the actual ray paths of direct P, pP, S2P and pS, direct S, and S2pS signals in Fig. 4.13, respectively. (a): Kernel computed from the direct P signal – 3D slices view; (b): Kernel computed from the direct P signal; (c): Kernel computed from the pP signal; (d): Kernel computed from the S2P and pS signals; (e): Kernel computed from the direct S and S2pS signals. 87
4.15 For $\alpha = 10\%$ and $\beta = 1/450$ m^{-2} in eq. (4.6), the subfigures (a) and (b) are the first two iterations using travel-time tomography, and the subfigure (c) is the third iteration using FWI which removes the artifacts in travel-time tomography and is close to the true velocity profile given in Fig. 4.12(b). 88
4.16 Two successive high-frequency ($\omega \gg 1$) seismic signals received at one station within the time period $[0, T]$. The distance between f and g is d, which implies the corresponding Wasserstein-2 metric is equal to d. . . . 88

5.1 Meta-architecture of a two-layered UNet, GeoDUDe-2, with Transfer Branch used in deep learning algorithms. For 2D problems the input is upsampled along the receiver axis by deconvolutions in the Transfer Branch. UNet's have "rungs" that connects the encoder and decoder branches. In this way, the network can incorporate both low and high resolution data. . . 96

5.2 The type of blocks used in GeoSeg for this paper: (a) Block compositions of a basic convolutional layer using a bottleneck convolution to expand the filter channels before the full convolution; (b) a corresponding dense block. Each layer of the block recieves input from all previous layers allowing information to flow through the whole block. 97
5.3 The locations of source and receivers, and the generated synthetic P-wave seismograms for the 1D interface problem. We take $k = 128$ for generating the synthetic data. (a) The source is located at (.5,.5,.5) km as a star and the 3 receivers are located on the surface. The interface presented is at a depth of 2 km. (b) A visualization of typical data point, which is a collection of 3 seismograms from the forward simulation using the FGA. . 100
5.4 1D interface predicted by GeoDUDe-3 using P-wave data. Each column of pixels represents a sample. The value of each pixel describes whether the material at the depth corresponding to that pixel's column belongs to either the high or low wavespeed region. 101
5.5 P-wave confidence distribution comparison produced by GeoDUDe-3 for 1D interface problem. Regions of low confidence correspond to areas where an interface is likely. 101
5.6 The locations of source and receivers, and the generated synthetic P- and S-wave seismograms for the 1D interface problem. We take $k = 32$ for generating the synthetic data. (a) The source is located at (.5,.5,.5) as a star and the 3 receivers are located on the surface. The interface presented is at a depth of 2 km. (b) A visualization of typical data point, which is a collection of 3 seismograms from the forward simulation using the FGA. 103
5.7 PS-wave training results for 1D interface problem, with synthetic data generated for $k = 32$ in (5.2): The evaluation data set for this figure only contains data generated by the FGA. 104
5.8 1D interface predicted by GeoDUDe-2 using P,S-wave data. Each column of pixels represents a sample. The value of each pixel describes whether the material at the depth corresponding to that pixel's column belongs to either the high or low wavespeed region. 104
5.9 P,S-wave heat-map distribution comparison produced by GeoDUDe-2 for 1D interface problem. Regions of low confidence correspond to areas where an interface is likely. 105
5.10 The locations of source and receivers, and the generated synthetic P- and S-wave seismograms for the three-layered media model. We take $k = 32$ for generating the synthetic data. (a) The source is located at $(.5, 1, .5)$ km as a star, the 32 receivers are located on the surface on the plane $y = 1$ km, and the interfaces presented are at a depth of 1.5 km and 2 km. (b) A visualization of typical data point, which is a collection of 32 seismograms from the forward simulation using the FGA. 106

5.11 Predictions for three-layered media by GeoDUDe-3: Each column of pixels represents a sample. The value of each pixel describes whether the material at the depth corresponding to that pixel's column belongs to either the high or low wavespeed region. There is a slight loss of confidence for the network detecting the lower interface. 106
5.12 Confidence map for three-layered media model produced by GeoDUDe-3. Regions of low confidence correspond to areas where an interface is likely. 107
5.13 The locations of source and receivers, and generated synthetic P- and S-wave seismograms for the 2D pocket model. We take $k = 32$ for generating the synthetic data. (a) The source is located at $(.5, 1, 1.5)$ km as a star and the 3 receivers are located on the surface on the plane $y = 1$ km. The interfaces are fixed at a depth of 1 km and 2.5 km. (b) A visualization of typical data point, which is a collection of 32 seismograms from the forward simulation using the FGA. 108
5.14 Visualization of network input using normalized displacement data for 2D pocket model. 109
5.15 Visualization of network input as image for 2D pocket model. Each color channel (inverse RGB) represents a coordinate of the displacement. . . . 109
5.16 2D pocket results predicted by GeoDUDe-4, with a typical data point chosen for visualization. The pocket is recovered with the networks confidence wavering on the boundary of the pocket. 109
5.17 Comparison of seismograms with noise and no noise for the 2D pocket model. (a) Seismogram with no noise. (b) Additive Gaussian white noise at 1% of max $|\mathbf{u}|$. This shows that 1% of the maximum recorded displacement is enough to mask the reflected data from the pocket. 110
5.18 Visualization of IOUs by GeoDUDe-4 for the 2D pocket model. Results taken from network trained with noise. Data is augmented with noise with a noise strength of 50%. (a) ground truth for comparision. (b) IOU=0.1403. (c) IOU= 0.2052. For each displayed results, the network was able to detect the location of the pocket. With additional noise the network is unable to resolve the geometry. 112
5.19 Network with trained without noise, 1000 data points are plotted in each Histogram. Subfigures (a), (b), (c) show the IOU metric with no noise, 10% noise strength, and 50% noise strength respectively. 112
5.20 Performance of the Network GeoDUDe-4 trained with noise strength at 20% of the average max displacement of the reflected wave for the 2D pocket model. 1000 data points are plotted in each Histogram. Subfigures (a), (b), (c) show the IOU metric, with no noise, 20% noise strength, and 50% noise strength, respectively. 113

List of Tables

5.1 P,S-Data Network Comparisons for 1D interface problem. The columns represent evaluation accuracy, training accuracy, and evaluation accuracy tested by SEM synthetic data. 103
5.2 IOU Scores for GeoDUDe-4 trained with and without noise for the 2D pocket model. 111

List of Symbols

$t \in [0,T]$	Temporal variable.
$\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$	Spatial variables.
$\mathbf{q} \in \mathbb{R}^d$	Position variable in the phase space.
$\mathbf{p} \in \mathbb{R}^d$	Momentum variable in the phase space.
$\mathbf{u} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$	Displacement vector.
$\mathcal{S}(\Omega)$	Schwartz class functions on a domain Ω.
$C^\infty(\Omega)$	Smooth functions on Ω.
$C_c^\infty(\Omega)$	Smooth functions with compact support on Ω.
$\rho : \mathbb{R}^d \to \mathbb{R}$	Mass density.
$\lambda : \mathbb{R}^d \to \mathbb{R}$	First Lamé parameter.
$\mu : \mathbb{R}^d \to \mathbb{R}$	Second Lamé parameter, shear modulus.
$\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$	External force
$c_\mathrm{p}, c_\mathrm{s} : \mathbb{R}^d \to \mathbb{R}$	P,S wave velocities, respectively.
$H : \mathbb{R}^{2d} \to \mathbb{R}$	Hamiltonian.
$K_\delta \subset \mathbb{R}^{2d}$	Compact set in the phase space with diameter with diameter $1/\delta$ and $\lvert\mathbf{p}\rvert > \delta$.
$\kappa : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$	Canonical transform
$\mathcal{F}^\epsilon, (\mathcal{F}^\epsilon)^*$	FBI Transform and its inverse.
$\mathcal{L} = c_\mathrm{p}^2 \mathrm{grad}(\mathrm{div}(\cdot)) - c_\mathrm{s}^2 \mathrm{curl}(\mathrm{curl}(\cdot))$	Spacial differential operator for elastic waves
$Z : \mathbb{R}^{2d} \to \mathbb{R}^{2d} \times \mathbb{R}^{2d}$	Complex phase space derivative
$\mathbf{u}_{\mathrm{F},0} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$	First order FGA.
$\mathbf{u}_{\mathrm{F},1} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$	Second order FGA.
$G^\epsilon \mathbb{R}^2 d \to \mathbb{R}$	Gaussian wave packet.
$\mathbf{u}_0, \mathbf{u}_1 : \mathbb{R}^d \to \mathbb{R}$	Initial displacement and velocity.
Id_n	$n \times n$ Identity matrix.

Chapter 1

Introduction

1.1 Permissions and Attributions

a. The content in chapter 2.3, 3.2, 4.2 and 4.3 is the result of a collaboration with Lihui Chai, Ping Tong and Xu Yang and previously appeared in Geophysical Journal International [35].

b. The content of chapter 5 is the result of a collaboration with Jay Roberts, Kyle Mylonakis and Xu Yang, and has been submitted to Geophysical Journal International [34].

All content reproduced here is with permission from collaborators.

1.2 Governing Equation

From Newton's second law,

$$\rho\partial_t^2\mathbf{u} = \mathbf{F} + \nabla\cdot\boldsymbol{\sigma}, \tag{1.1}$$

where $\rho : \mathbb{R}^n \to \mathbb{R}$ is the mass density, $\boldsymbol{\sigma} : \mathbb{R}^3 \to \mathbb{R}^3$ is Cauchy stress tensor, $\mathbf{u} : \mathbb{R}^3 \to \mathbb{R}^3$ is displacement, $\mathbf{F}$ is an external Force and $\nabla\cdot$ being the vector divergence operator. Using Hooke's law with the stiffness tensor $\mathbf{C}$ and linear strain-displacement $\boldsymbol{\epsilon}$, we have the constitutive equations; $\boldsymbol{\sigma} = \mathbf{C} : \boldsymbol{\epsilon}$, with $:$ being a contraction and $2\boldsymbol{\epsilon} = \nabla\mathbf{u} + \nabla\mathbf{u}^T$ Assuming linear stress-strain, plugging in and simplifying the differential operators the equation becomes:

$$\rho\partial_t^2 = \mathbf{F} + \nabla\Big(\lambda\,\nabla\cdot\mathbf{u} + \mu\big(\nabla\mathbf{u} + (\nabla\mathbf{u})^T\big)\Big) \tag{1.2}$$

with λ the first Lamé parameter, μ is the shear modulus. Expanded (1.2) is:

$$\rho\mathbf{u}_{tt} = \mathbf{F} + \nabla\lambda(\nabla\cdot\mathbf{u}) + \nabla\mu\cdot(\nabla\mathbf{u} + (\nabla\mathbf{u})^T) + (\lambda+\mu)\nabla(\nabla\cdot\mathbf{u}) + \mu\boldsymbol{\Delta}\mathbf{u}. \tag{1.3}$$

The gradient of the Lamé parameters $\nabla\lambda, \nabla\mu$ are typically small [109, 48]. In terms of the asymptotic expansion to come, they will not contribute to the leading order terms, this will be further remarked upon in Chapter 2.1. It is noted now that, in terms of the asymptotic expansion, this implies that an isotropy is a first order approximation for an anisotropy.

The isotropic elastic wave equation is;

$$\rho\partial_t^2\mathbf{u} = \mathbf{F} + (\lambda+\mu)\nabla(\nabla\cdot\mathbf{u}) + \mu\boldsymbol{\Delta}\mathbf{u}. \tag{1.4}$$

Using the an identity from vector calculus, $\boldsymbol{\Delta} = \nabla(\nabla\cdot) - \nabla\times\nabla\times$,equation (1.4) can also be written as

$$\rho\partial_t^2\mathbf{u} = \mathbf{F} + (\lambda+2\mu)\nabla(\nabla\cdot\mathbf{u}) - \mu\nabla\times\nabla\times\mathbf{u} \tag{1.5}$$

Equation (1.5) can be written in terms of a Helmholtz decomposition by separating the

wave fields into a curl and div free parts. This decomposition shows the potentials for the two types of body waves. Applying the div operator, letting $\nabla \cdot u = \phi$.

$$\rho \partial_t^2 \phi = \nabla \cdot \mathbf{F} + (\lambda + 2\mu)\Delta\phi \tag{1.6}$$

which is a scalar wave equation with external force $F_{\mathrm{p}} = \nabla \cdot \mathbf{F}$ and velocity

$$c_{\mathrm{p}} = \sqrt{\frac{\lambda + 2\mu}{\rho}} \tag{1.7}$$

Equation (1.6) represents primary waves; these are compressional waves as displacement occurs in the direction of the propagation. Applying the curl operator,

$$\rho \partial_t^2 (\nabla \times u_i) = \nabla \times \mathbf{F} + (\lambda + 2\mu)\nabla \times \nabla\phi + \mu\Delta(\nabla \times u_i) \tag{1.8}$$

Let $\nabla \times \mathbf{u} = \psi$, as $\nabla \times \nabla\phi = 0$, equation (1.5) becomes a vector wave equation

$$\rho \partial_t^2 \psi = \nabla \times \mathbf{F} + \Delta\psi, \tag{1.9}$$

with external force $\mathbf{F}_{\mathrm{s}} = \nabla \times \mathbf{F}$ and velocity

$$c_{\mathrm{s}} = \sqrt{\frac{\mu}{\rho}}. \tag{1.10}$$

Equation (1.9) represents secondary waves; these are shear waves as displacement is traverse to the direction of the propagation. The Helmholtz decomposition is realized as:

$$\mathbf{u} = \nabla\Phi + \nabla \times \Psi \tag{1.11}$$

with $\phi = \Delta\Phi$, and $\psi = -\Delta\Psi$.

1.3 WKBJ Approximation

For motivation into splitting the wave fields for the FGA formulation I turn to the WKBJ approximation. The WKBJ method is a semiclassical method, most commonly used in quantum mechanics and geometric optics. The solution ansatz is an exponential function; semiclassically expanded, usually with either a slowly changing amplitude or the phase [28]. For simplicity consider the isotropic case; meaning the Lamé parameters are constant. The ansatz is:

$$\mathbf{u}(t,\mathbf{x}) = \mathbf{a}(t,\mathbf{x})e^{\mathrm{i}S(t,\mathbf{x})/\epsilon}, \qquad \mathbf{a}(t,\mathbf{x}) = a_j(t,\mathbf{x}) \tag{1.12}$$

for $j = 1..3$, with ϵ as a small parameter. Computing the necessary components to plug in:

$$\begin{aligned}
\mathbf{u}_{tt} &= \left(\mathbf{a}_{tt} + 2\frac{\mathrm{i}}{\epsilon}\mathbf{a}_t S_t + \frac{\mathrm{i}}{\epsilon}\mathbf{a}S_{tt} - \frac{\mathbf{a}S_t^2}{\epsilon^2}\right) e^{\mathrm{i}\mathbf{S}(t,x)/\epsilon} \\
\nabla(\nabla\cdot\mathbf{u}) &= \left(\nabla(\nabla\cdot\mathbf{a}) + \frac{\mathrm{i}}{\epsilon}\Big((\nabla\cdot\mathbf{a})\nabla S + \nabla(\mathbf{a}\cdot\nabla S)\Big) - \frac{1}{\epsilon^2}(\mathbf{a}\cdot\nabla S)\nabla S\right) e^{\mathrm{i}\mathbf{S}(t,x)/\epsilon} \\
\nabla^2\mathbf{u} &= \left(\nabla^2\mathbf{a} + \frac{\mathrm{i}}{\epsilon}\Big(2\nabla\mathbf{a}\nabla S + \mathbf{a}\nabla^2 S\Big) - \frac{1}{\epsilon^2}\mathbf{a}|\nabla S|^2\right) e^{\mathrm{i}\mathbf{S}(t,x)/\epsilon} \\
\nabla\times\nabla\times\mathbf{u} &= \left(\nabla\times\nabla\times\mathbf{a} + \frac{\mathrm{i}}{\epsilon}\Big(\nabla S\times(\nabla\times\mathbf{a})\Big) - \frac{1}{\epsilon^2}\nabla S\times(\nabla S\times\mathbf{a})\right) e^{\mathrm{i}\mathbf{S}(t,x)/\epsilon} \\
&\quad + \frac{\mathrm{i}}{\epsilon}\nabla\times(\nabla S\times\mathbf{a})e^{\mathrm{i}\mathbf{S}(t,x)/\epsilon}
\end{aligned} \tag{1.13}$$

Plugging this into the equation (1.5), we arrive at

$$\rho\left(\mathbf{a}_{tt} + 2\frac{\mathrm{i}}{\epsilon}\mathbf{a}_t S_t + \frac{\mathrm{i}}{\epsilon}\mathbf{a}S_{tt} - \frac{\mathbf{a}S_t^2}{\epsilon^2}\right) - \mathbf{F}e^{-\mathrm{i}S(t,x)/\epsilon} \tag{1.14}$$

$$= (\lambda+\mu)\left(\nabla(\nabla\cdot\mathbf{a}) + \frac{\mathrm{i}}{\epsilon}\Big((\nabla\cdot\mathbf{a})\nabla S + \nabla(\mathbf{a}\cdot\nabla S)\Big) - \frac{1}{\epsilon^2}(\mathbf{a}\cdot\nabla S)\nabla S\right) \tag{1.15}$$

$$+ \mu\left(\nabla^2\mathbf{a} + \frac{\mathrm{i}}{\epsilon}\Big(2\nabla\mathbf{a}\nabla S + \mathbf{a}\nabla^2 S\Big) - \frac{1}{\epsilon^2}\mathbf{a}|\nabla S|^2\right) \tag{1.16}$$

Leading order term in the WKBJ approximation

Let $\mathbf{a} = \epsilon^n \mathbf{a}_n$ with sum over n be the asymptotic expansion. The leading order term is of $\mathcal{O}(\epsilon^{-2})$ and gives the equation:

$$\rho \mathbf{a}_0 S_t^2 = (\lambda + \mu)(\mathbf{a}_0 \cdot \nabla S)\nabla S + \mu \mathbf{a}_0 |\nabla S|^2 \tag{1.17}$$

Taking the inner product with ∇S, we have

$$\rho \mathbf{a}_0 S_t^2 \cdot \nabla S \quad = \quad (\lambda + \mu)(\mathbf{a}_0 \cdot \nabla S)\nabla S \cdot \nabla S + \mu |\nabla S|^2 \mathbf{a}_0 \cdot \nabla S$$

Which gives

$$(\mathbf{a}_0 \cdot \nabla S)\left(\rho(x) S_t^2 - (\lambda + 2\mu)|\nabla S|^2\right) = 0 \tag{1.18}$$

If $\mathbf{a}_0 \cdot \nabla S = 0$, then the $\mathcal{O}(\epsilon^{-2})$ reduces to an Eikonal equation,

$$\rho(x) S_t^2 = \mu |\nabla S|^2 \qquad \Rightarrow \qquad S_t = \pm c_{\mathrm{s}} \left|\nabla S\right|^2 \tag{1.19}$$

where c_{s} is the S-wave velocity in (1.10). The Hamiltonian is

$$H_{\mathrm{s},\pm}(\mathbf{Q}_\pm, \mathbf{P}_\pm) = \pm c_{\mathrm{s}}(\mathbf{Q}_\pm)|\mathbf{P}_\pm| \tag{1.20}$$

with flow

$$\frac{\mathrm{d}\mathbf{Q}_\pm}{\mathrm{d}t} = \pm c_{\mathrm{s}}(\mathbf{Q}_\pm)\frac{\mathbf{P}_\pm}{|\mathbf{P}_\pm|}, \qquad \frac{\mathrm{d}\mathbf{P}_\pm}{\mathrm{d}t} = \mp \nabla_{\mathbf{Q}} c_{\mathrm{s}}(\mathbf{Q}_\pm)|\mathbf{P}| \tag{1.21}$$

The Hamiltonian (1.20) and corresponding system (1.21) give the dynamics for the secondary waves. Otherwise, returning to (1.18),

$$\rho(x) S_t^2 - (\lambda + 2\mu)|\nabla S|^2 = 0. \tag{1.22}$$

This leads to a similar result for primary waves;

$$\rho(x)S_t^2 = (\lambda + 2\mu)|\nabla S|^2 \qquad \Rightarrow \qquad S_t = \pm c_{\mathrm{p}}\big|\nabla S\big|. \tag{1.23}$$

Again, the Hamiltonian is

$$H_{p,\pm}(\mathbf{Q}_\pm, \mathbf{P}_\pm) = \pm c_{\mathrm{p}}(\mathbf{Q}_\pm)|\mathbf{P}_\pm| \tag{1.24}$$

with flow

$$\frac{\mathrm{d}\mathbf{Q}_\pm}{\mathrm{d}t} = \pm c_{\mathrm{p}}(\mathbf{Q}_\pm)\frac{\mathbf{P}_\pm}{|\mathbf{P}_\pm|}, \qquad \frac{\mathrm{d}\mathbf{P}_\pm}{\mathrm{d}t} = \mp \nabla_{\mathbf{Q}} c_{\mathrm{p}}(\mathbf{Q}_\pm)|\mathbf{P}| \tag{1.25}$$

Second term in the WKB Approximation

The next term in the WKB Approximation is $\mathcal{O}(\epsilon^{-1})$. The equation from this term is;

$$\begin{aligned}\rho(x)\left(2(\mathbf{a}_0)_t S_t + \mathbf{a}_0 S_{tt} + \mathrm{i}\mathbf{a}_1 S_t^2\right) = \\ (\lambda+\mu)\big((\nabla\cdot\mathbf{a}_0)\nabla S + \nabla(\mathbf{a}_0\cdot\nabla S) \\ + \mathrm{i}(\mathbf{a}_1\cdot\nabla S)\nabla S\big) + \mu(2\nabla\mathbf{a}_0\nabla S + \mathbf{a}_0\nabla^2 S + \mathrm{i}\mathbf{a}_1|\nabla S|^2).\end{aligned} \tag{1.26}$$

Consider the S-waves, $\mathbf{a}\cdot\nabla S = 0$. Taking inner product with $\mathbf{a}_0$ in the order $\mathcal{O}(\epsilon^{-1})$ reduces to

$$\rho|\mathbf{a}_0|_t^2 S_t + \rho|\mathbf{a}_0|^2 S_{tt} = \mu\,\nabla\cdot(|\mathbf{a}_0|^2\nabla S), \tag{1.27}$$

which is a transport equation. Consider the P-waves and substituting equation (1.23) into equation (1.17) gives

$$\mathbf{a}_0 = \frac{\mathbf{a}_0\cdot\nabla S}{|\nabla S|^2}\nabla S. \tag{1.28}$$

This tells us that $\mathbf{a}_0$ is parallel to the direction of the propagation ∇S, $\mathbf{a}_0 \times \nabla S = 0$. Plugging equation (1.28) in to equation (1.26) with simplification gives;

$$\rho|\mathbf{a}_0|_t^2 S_t + \rho|\mathbf{a}_0|^2 S_{tt} = (\lambda + 2\mu)\big((\nabla\cdot \mathbf{a}_0)(\mathbf{a}_0 \cdot \nabla S) + \nabla(\mathbf{a}_0 \cdot \nabla S) \cdot \mathbf{a}_0\big). \tag{1.29}$$

Which can be recast as another transport equation.

1.4 Introduction to the FGA

For modeling high frequency waves seismic waves, mesh based methods such as finite difference methods [94, 95, 57, 92, 84, 52, 29], finite element methods [7, 80, 100], spectral methods [88, 18, 89, 55] and pseudo-spectral methods [27, 98, 17] have the same obstacle to overcome. With a large domain size, on the order of tens or hundreds of kilometers, and with a dominant frequency for typical earthquake around 5 Hz [65], this leads to demanding, and at times, unaffordable computational cost. Two other well known methods for solving high frequency wave propagation are geometric optics and Gaussian beam method. Geometric optics solve Eikonal and transport equations [23, 79]. This makes the choice of mesh size frequency-independent; however, Eikonal equation can develop singularities which is a problematic at caustics [15]. The Gaussian beam method has the advantage of overcome caustic problems. For Gaussian beams, the rays determined by the Hamiltonian system related to the Eikonal equation serve as the centers of the beams. As opposed to geometric rays, the Gaussian beam method allows the phase function to be complex off its center. The imaginary part of phase function is positive, which makes the solution decay exponentially away from the center. The draw back is that the accuracy of the Gaussian beams relies on a Taylor expansion which determines by the width of beams, thus the error of the approximation increases when beams spread.

This is known as beam spreading. [10, 42, 87, 72, 64, 58]. When the imaginary part of the phase function becomes small enough, the Gaussian is not localized so the solution of the wave equation spreads.

The FGA initially comes from quantum chemistry. To fix the problem of beam spreading frozen Gaussian wave packets were introduced by Heller in 1981 [38] for a semi-classical method. Shortly after; in 1984, the Herman-Kluk Propagator [41] was derived from a class of Fourier integral operators which converge to the unitary group of the Schödingers equation in the semi-classical regime. Ten year later, in 1994, Kay numerically verified the Herman-Kluk propagator has especially high accuracy and rapid convergence [49]. Kay derives the FGA and provided numerical evidence for computing time-dependent wave functions. In 2009 Swart and Rousse, give an analytical proof of the convergence of the Herman-Kluk propagator [85], using the FBI transform. Then multiple works regarding high frequency waves for hyperbolic systems [58, 59, 60], with applications in seismology, [101, 11, 12] and more recent works with Schrödinger equation [19, 20].

1.4.1 Notation.

In this section the necessary tools for theoretical discussion on the FGA are defined. First define the closed set $K_\delta \subset \mathbb{R}^{2d}$

$$K_\delta = \Big\{(\mathbf{q},\mathbf{p}) \in \mathbb{R}^{2d} : |\mathbf{q}| \leq 1/\delta, \quad \delta \leq |\mathbf{p}| \leq 1/\delta\Big\} \tag{1.30}$$

For all practical purpose, this set K_δ is bounding the position and the magnitude for the direction of propagation of the wave packets. The upper bound condition on $|\mathbf{q}|$ and $|\mathbf{p}|$ is reasonable as any computational domain will be a finite domain. $\mathbf{p}$ bounded away from zero is reasonable as if $\mathbf{p} = 0$ the wave packet does not propagate and the Hamiltonian

system is degenerate i.e., $H = 0$.

We use the notation $\mathcal{O}(\epsilon^\infty)$: $A^\epsilon = \mathcal{O}(\epsilon^\infty)$ meaning for any $k \in \mathbb{N}$

$$\lim_{\epsilon \to 0} \epsilon^{-k} |A^\epsilon| = 0. \tag{1.31}$$

Notation C will be used as a general constant, that can vary from line to line. For the theoretical discussion the value will not be important, but rather that it is finite. Subscripts will be used to denote constant dependence, e.g. C_T, is a constant that depends on the parameter T. The usual notation of $\mathcal{S}$, C^∞ and C_c^∞ for the Schwartz class, smooth and compacted supported smooth functions respectively. For generality $\mathbb{R}^d$ will be used for a d-dimensional Euclidean space; however, for the actually equation and computations we set $d = 3$ as we deal the usual differential operators on $\mathbb{R}^3$.

1.4.2 The FBI Transform

In this section the integral transform for the FGA is defined. Also necessary propositions for the proof of convergence are cited. For $(\mathbf{q}, \mathbf{p}) \in \mathbb{R}^{2d}$, define $\phi^\epsilon_{\mathbf{q},\mathbf{p}}$ as

$$\phi^\epsilon_{\mathbf{q},\mathbf{p}}(\mathbf{x}) = (-2\pi\epsilon)^{d/2} \exp\Big(\mathrm{i}\mathbf{p} \cdot (\mathbf{x} - \mathbf{q})/\epsilon - |\mathbf{x} - \mathbf{q}|^2/(2\epsilon)\Big) \tag{1.32}$$

The Fourier–Bros–Iagolnitzer (FBI) transform on $\mathcal{S}(\mathbb{R}^d)$ is defined as;

$$(\mathcal{F}^\epsilon f)(\mathbf{q}, \mathbf{p}) = \pi\epsilon^{-d/4} \langle \phi^\epsilon_{\mathbf{q},\mathbf{p}}, f \rangle \tag{1.33}$$

$$= 2^{-d/2} (\pi\epsilon)^{-3d/4} \int_{\mathbb{R}^d} \exp\Big(\mathrm{i}\mathbf{p} \cdot (\mathbf{x} - \mathbf{q})/\epsilon - |\mathbf{x} - \mathbf{q}|^2/(2\epsilon)\Big) f(\mathbf{x}) \, \mathrm{d}\mathbf{x} \tag{1.34}$$

With inverse transform $\left(\mathcal{F}^\epsilon\right)^*$ defined on $\mathcal{S}(\mathbb{R}^{2d})$ given by

$$\Big(\left(\mathcal{F}^\epsilon\right)^* f\Big)(\mathbf{x}) = 2^{-d/2}(\pi\epsilon)^{-3d/4}\int_{\mathbb{R}^{2d}} \exp\Big(\mathrm{i}\mathbf{p}\cdot(\mathbf{x}-\mathbf{q})/\epsilon - |\mathbf{x}-\mathbf{q}|^2/(2\epsilon)\Big) g(\mathbf{q},\mathbf{p})\, \mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}. \tag{1.35}$$

The FBI transform is a localized Fourier transform. The original motivation of the FBI transform from Bros and Iagolnitzer was to analyze the analytic wave front set of distributions that appear in quantum field scattering theory [45]. The FBI can be thought of as a Fourier transform with a convolution of a Gaussian. This convolution operation mollifies the singularities of distributions. The FBI transform and its inverse from the integral transform for the FGA. Note when $\alpha = 0$, (1.32) coincides with the Fourier transform. As initially the transform lifts the function to the phase space, the FBI transform and its inverse are not commutative as operators. Projecting from the phase space to physical space and then lift back to the phase space leads to a "loss of information".

Proposition 1.4.1 *For the Schwartz class, the FBI transform is an isometry on $\mathbb{R}^d$, i.e., for any $f \in S(\mathbb{R}^d)$*

$$\|\mathcal{F}^\epsilon f\|_{\mathrm{L}^{2d}} = \|f\|_{\mathrm{L}^{2d}} \tag{1.36}$$

Furthermore; $\left(\mathcal{F}^\epsilon\right)^*\mathcal{F}^\epsilon = \mathrm{Id}_{\mathrm{L}^2(\mathbb{R}^d)}$. Meaning the domain of $\mathcal{F}^\epsilon$ and $\left(\mathcal{F}^\epsilon\right)^*$ can be extended to $\mathrm{L}^2(\mathbb{R}^d)$ and $\mathrm{L}^2(\mathbb{R}^{2d})$ respectively.

Proof: This is a standard results in micro-local analysis, a proof can be found here [6]. ■

Definition 1.4.2 *Let $\{u^\epsilon\} \subset \mathrm{L}^2(\mathbb{R}^d)$ be a family of functions that is uniformly bounded. Given $\delta > 0$, $\{\mathbf{u}^\epsilon\}$ is asymptotically high frequency with cut off δ,* if

$$\int_{\mathbb{R}^{2d}\setminus K_\delta} |(\mathcal{F}^\epsilon u^\epsilon)(\mathbf{q},\mathbf{p})|^2\, \mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} = \mathcal{O}(\epsilon^\infty) \tag{1.37}$$

as $\epsilon \to 0$.

Definition 1.4.3 *For $M \in \mathrm{L}^\infty(\mathbb{R}^{2d}; \mathbb{C}^{N\times N})$ and a Schwartz function $\mathbf{u} \in \mathcal{S}(\mathbb{R}^d; \mathbb{C}^N)$ for each $n = 1, \ldots, N$ define the Fourier integral operator $(\mathcal{I}_n^\epsilon(t, M)\mathbf{u})(\mathbf{x})$ as*

$$(\mathcal{I}_n^\epsilon(t, M)\mathbf{u})(\mathbf{x}) = (2\pi\epsilon)^{-3d/2} \int_{\mathbb{R}^{3d}} e^{\mathrm{i}\phi_n(t,\mathbf{x},\mathbf{y},\mathbf{p},\mathbf{q})/\epsilon} M(\mathbf{q}, \mathbf{p})\mathbf{u}(\mathbf{y})\, \mathrm{d}\mathbf{q}\, \mathrm{d}\mathbf{p}\, \mathrm{d}\mathbf{y} \tag{1.38}$$

With phase function

$$\begin{aligned}\phi_n(t, \mathbf{x}, \mathbf{y}, \mathbf{p}, \mathbf{q}) = \frac{\mathrm{i}}{2}|\mathbf{y} - \mathbf{q}|^2 - \mathbf{p} \cdot (\mathbf{y} - \mathbf{q}) \\ + \frac{\mathrm{i}}{2}|\mathbf{x} - \mathbf{Q}_n(t, \mathbf{q}, \mathbf{p})|^2 + \mathbf{P}_n(t, \mathbf{q}, \mathbf{p}) \cdot (\mathbf{x} - \mathbf{Q}_n(t, \mathbf{q}, \mathbf{p})).\end{aligned} \tag{1.39}$$

Equation (1.38) is the integral operator for the FGA defined in the next chapter.

Proposition 1.4.4 *If $M \in \mathrm{L}^\infty(\mathbb{R}^{2d}; \mathbb{C}^{N\times N})$, for any t and each $n = 1, \ldots, N$, $\mathcal{I}_n^\epsilon(t, M)$ can be extended to a bounded linear operator on $\mathrm{L}^2(\mathbb{R}^d; \mathbb{C}^N)$ with bound*

$$\|\mathcal{I}_n^\epsilon(t, M)\|_{\mathcal{L}(\mathrm{L}^2(\mathbb{R}^d;\mathbb{C}^N))} \leq 2^{-d/2}\|M\|_{\mathrm{L}^\infty(\mathbb{R}^d;\mathbb{C}^N))} \tag{1.40}$$

This is Proposition 3.7 in [60], a more general version is also proved [85], see Theorem 2.

Chapter 2

Frozen Gaussian Approximation

This chapter covers the construction of the FGA and the proof of its convergence. First the FGA is generalized to solve 3-D elastic wave propagation, the derivation of FGA formulation requires to do the asymptotic expansion in an integral form so that one is able to perform integration by parts to eliminate the extra constraints yielded by direct asymptotic expansion. In order to derive the evolution equations a projection onto a specific orthonormal basis is used. Proof of convergence and asymptotic equivalence to the eigenfunction decomposition are shown in the last subsection.

2.1 Preliminaries

Let $\mathbf{u} : \mathbb{R}^+ \times \mathbb{R}^3 \to \mathbb{R}^3$ and $c_{\rm p}, c_{\rm s} : \mathbb{R}^3 \to \mathbb{R}$, denote the operator $\mathcal{L}\mathbf{u} = c_{\rm p}^2 \nabla\big(\nabla \cdot \mathbf{u}\big) - c_{\rm s}^2 \, \nabla\times \big(\nabla\times \mathbf{u}\big)$, with the differential operators taken in the spacial variables and

$$c_{\rm p}^2(\mathbf{x}) = \frac{\lambda(\mathbf{x}) + 2\mu(\mathbf{x})}{\rho(\mathbf{x})} \qquad c_{\rm s}^2(\mathbf{x}) = \frac{\mu(\mathbf{x})}{\rho(\mathbf{x})}. \tag{2.1}$$

We are interested in the Cauchy problem with high frequency initial datum;

$$\begin{cases} \partial_t^2 \mathbf{u} - \mathcal{L}\mathbf{u} = 0 \\ u(0,\mathbf{x}) = \mathbf{u}_0^\epsilon, \qquad \mathbf{x} \in \mathbb{R}^d \\ u(t,\mathbf{x}) = \mathbf{u}_1^\epsilon, \qquad \mathbf{x} \in \mathbb{R}^d \end{cases} \tag{2.2}$$

With this formulation is is noted that $\epsilon \mathbf{u}_1^\epsilon$ and $\mathbf{u}_0^\epsilon$ are on the same order of magnitude. As previously mentioned, it is also remarked that as $\nabla\lambda, \nabla\mu << 1$, it will be assumed that λ and μ are constant. As it will be seen later, this assumption will not change the first order FGA approximation. Define the following quantities

$$\Theta(t,\mathbf{x}) = \nabla\cdot \mathbf{u}(t,\mathbf{x}), \qquad \Psi(t,\mathbf{x}) = \nabla\times \mathbf{u}(t,\mathbf{x}), \qquad \mathbf{v}(t,\mathbf{x}) = \partial_t \mathbf{u}(t,\mathbf{x}). \tag{2.3}$$

With $\mathbf{v} = (v_1, v_2, v_3)^{\mathrm{T}}$ and $\Psi = (\Psi_1, \Psi_2, \Psi_3)^{\mathrm{T}}$, let $X = (v_1, v_2, v_3, \Theta, \Psi_1, \Psi_2, \Psi_3)^{\mathrm{T}}$. Eq. (2.2) can be written as a matrix system; in terms of the auxiliary variables (2.3),

$$\partial_t X = M_x \partial_x X + M_y \partial_y X + M_z \partial_z X \tag{2.4}$$

Eq. (2.2) can be written as a matrix system; in terms of the axillary variables (2.3),

$$\partial_t X = M_x \partial_x X + M_y \partial_y X + M_z \partial_z X. \tag{2.5}$$

Using sparse notation; e.g. $M_{ij} = v$ is denoted (i, j, v), the M_x, M_y, M_z are as follows:

$$\begin{aligned} M_x &: (1,4,c_{\mathrm{p}}^2),(2,7,c_{\mathrm{s}}^2),(3,6,-c_{\mathrm{s}}^2),(4,1,1),(7,2,1),(6,3,-1), \\ M_y &: (1,7,-c_{\mathrm{s}}^2),(2,4,c_{\mathrm{p}}^2),(3,5,c_{\mathrm{s}}^2),(4,2,1),(5,3,1),(7,1,-1), \\ M_z &: (1,6,c_{\mathrm{s}}^2),(2,5,-c_{\mathrm{s}}^2),(3,4,c_{\mathrm{p}}^2),(4,3,1),(6,1,1),(5,2,-1). \end{aligned} \tag{2.6}$$

For example;

$$M_x = \begin{pmatrix} 0 & 0 & 0 & c_{\mathrm{p}}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & c_{\mathrm{s}}^2 \\ 0 & 0 & 0 & 0 & 0 & -c_{\mathrm{s}}^2 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.7}$$

The eigenvalues of $M_x + M_y + M_z$ are, $\pm c_{\mathrm{p}}, 0, \pm c_{\mathrm{s}}$, with $\pm c_{\mathrm{s}}$ each having a multiplicity of 2. Once $\mathbf{v}$ is known, solving for $\mathbf{u}$ is an easy enough task so it's not included in the matrix system. We note here that, Θ represents the potential for the P-waves and Ψ the potential of the S-waves.

Let M denote the operator $M = M_x\partial_x + M_y\partial_y + M_z\partial_z$, the solution to the system (2.5) can be written abstractly as a one parameter semi-group;

$$X(t) = \exp(tM)X(0) \tag{2.8}$$

With X(0) defined appropriately from the Cauchy problem in Eq. (2.2). The Eigenfunctions and left and right eigen vectors of the matrix system are as follows:

$$H_0(\mathbf{q},\mathbf{p}) = 0 \tag{2.9}$$
$$R_0 = (0,0,0,0,p_1,p_2,p_3)^{\mathrm{T}}$$
$$L_0 = \frac{1}{|p|^2}(0,0,0,0,p_1,p_2,p_3)^{\mathrm{T}}$$
$$H_{\mathrm{p}\pm}(\mathbf{q},\mathbf{p}) = \pm c_{\mathrm{p}}(\mathbf{q})|\mathbf{p}| \tag{2.10}$$
$$R_{\mathrm{p}\pm} = (p_1,p_2,p_3,H_{\mathrm{p}\pm}/c_{\mathrm{p}}^2,0,0,0)^{\mathrm{T}},$$
$$L_{\mathrm{p}\pm} = \frac{c_{\mathrm{p}}^2}{2H_{\mathrm{p}\pm}}(p_1,p_2,p_3,H_{\mathrm{p}\pm},0,0,0)^{\mathrm{T}}$$
$$H_{\mathrm{s}\pm}(\mathbf{q},\mathbf{p}) = \pm c_{\mathrm{s}}(\mathbf{q})|\mathbf{p}| \tag{2.11}$$
$$R_{\mathrm{p}\pm,1} = (-c_{\mathrm{s}}(\mathbf{q})p_1p_2, c_{\mathrm{s}}(\mathbf{q})(p_1^2+p_3^2), c_{\mathrm{s}}(\mathbf{q})p_2p_3, 0, -p_3H_{\mathrm{s}\pm}, 0, p_1H_{\mathrm{s}\pm})^{\mathrm{T}},$$
$$L_{\mathrm{s}\pm,1} = \frac{-1}{H_{\mathrm{s}\pm}p_1p_2}(p_2^2+p_3^2, -p_1p_2, -p_1p_3, 0, 0, p_3H_{\mathrm{s}\pm}, -p_2H_{\mathrm{s}\pm})^{\mathrm{T}}$$
$$H_{\mathrm{s}\pm}(\mathbf{q},\mathbf{p}) = \pm c_{\mathrm{s}}(\mathbf{q})|\mathbf{p}| \tag{2.12}$$
$$R_{\mathrm{p}\pm,2} = (c_{\mathrm{s}}(\mathbf{q})p_1p_3, c_{\mathrm{s}}(\mathbf{q})(p_2p_3), -c_{\mathrm{s}}(\mathbf{q})(p_1^2+p_2^2), 0, -p_2H_{\mathrm{s}\pm}, p_1H_{\mathrm{s}\pm}, 0)^{\mathrm{T}},$$
$$L_{\mathrm{s}\pm,2} = \frac{-1}{H_{\mathrm{s}\pm}p_2p_3}(p_2p_1, -p_1^2-p_3^2, p_3p_2, 0, p_3H_{\mathrm{s}\pm}, -p_1H_{\mathrm{s}\pm})^{\mathrm{T}}$$

Assuming $p_1, p_2, p_3 \neq 0$, normalization is given so that $L_m \cdot R_n = \delta_{n,m}$; otherwise, the eigenvectors can be adjusted accordingly. Equations (2.9)-(2.12) are used for the eigenvector decomposition form of the FGA; they will not be used for deriving the evolution equations via projection. Eqs. (2.9)-(2.12) are written for the sake of completeness. The Hamiltonian associated with Θ, Ψ

$$H(t,\mathbf{Q},\mathbf{P})_{\mathrm{p}\pm,\mathrm{s}\pm} = \pm c_{\mathrm{p,s}}(\mathbf{Q}_{\mathrm{p}\pm,\mathrm{s}\pm}(t,\mathbf{q},\mathbf{p}))\mathbf{P}_{\mathrm{p}\pm,\mathrm{s}\pm}(t,\mathbf{q},\mathbf{p}) \tag{2.13}$$

The corresponding flows are given by

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{Q}_{\mathrm{p}\pm,\mathrm{s}\pm}}{\mathrm{d}t}(t,\mathbf{q},\mathbf{p}) &= \pm c_{\mathrm{p,s}}(\mathbf{Q}_{\mathrm{p}\pm,\mathrm{s}\pm}(t,\mathbf{q},\mathbf{p}))\dfrac{\mathbf{P}_{\mathrm{p}\pm,\mathrm{s}\pm}(t,\mathbf{q},\mathbf{p})}{|\mathbf{P}_{\mathrm{p}\pm,\mathrm{s}\pm}(t,\mathbf{q},\mathbf{p})|} \\ \dfrac{\mathrm{d}\mathbf{P}_{\mathrm{p}\pm,\mathrm{s}\pm}}{\mathrm{d}t}(t,\mathbf{q},\mathbf{p}) &= \mp\partial_{\mathbf{Q}} c_{\mathrm{p,s}}(\mathbf{Q}_{\mathrm{p}\pm,\mathrm{s}\pm}(t,\mathbf{q},\mathbf{p}))|\mathbf{P}_{\mathrm{p}\pm,\mathrm{s}\pm}(t,\mathbf{q},\mathbf{p})| \end{cases} \tag{2.14}$$

with initial conditions

$$\mathbf{Q}_{\mathrm{p}\pm,\mathrm{s}\pm}(0,\mathbf{q},\mathbf{p}) = \mathbf{q} \text{ and } \mathbf{P}_{\mathrm{p}\pm,\mathrm{s}\pm}(0,\mathbf{q},\mathbf{p}) = \mathbf{p} \tag{2.15}$$

We remark that,

$$|\mathbf{p}\cdot\partial_{\mathbf{q}} H(t,\mathbf{q},\mathbf{p})| \lesssim |\mathbf{p}|^2 \qquad \text{and} \qquad |\mathbf{q}\cdot\partial_{\mathbf{p}} H(t,\mathbf{q},\mathbf{p})| \lesssim |\mathbf{q}|^2 \tag{2.16}$$

so assumption A in [60] is satisfied. This is equivalent to a global Lipschitz condition.

Proposition 2.1.1 *For $T > 0$ and $\delta > 0$, there is a constant δ_T such that*

$$(\mathbf{Q}_{\mathrm{p}\pm,\mathrm{s}\pm}(t,\mathbf{q},\mathbf{p}), \mathbf{P}_{\mathrm{p}\pm,\mathrm{s}\pm}(t,\mathbf{q},\mathbf{p})) \in K_{\delta_T} \tag{2.17}$$

Proof: This is proposition 3.1 in [60], the proof requires the bound in eq. (2.16) and Gronwall's inequailty. ■

For the following definition we omit the branch subscript.

Definition 2.1.2 *A map $\kappa_{\mathrm{p,s}} : (\boldsymbol{q},\boldsymbol{p}) \to \big(\boldsymbol{Q}_{\mathrm{p,s}}(\boldsymbol{q},\boldsymbol{p}), \boldsymbol{P}_{\mathrm{p,s}}(\boldsymbol{q},\boldsymbol{p})\big)$ is called canonical transformation if the associated Jacobian matrix is symplectic, i.e., for any $(\mathbf{q},\mathbf{p})$*

$$J_{\mathrm{p,s}}(\boldsymbol{q},\boldsymbol{p}) = \begin{pmatrix} (\partial_q \boldsymbol{Q}_{\mathrm{p,s}})^T(\boldsymbol{q},\boldsymbol{p}) & (\partial_p \boldsymbol{Q}_{\mathrm{p,s}})^T(\boldsymbol{q},\boldsymbol{p}) \\ (\partial_q \boldsymbol{P}_{\mathrm{p,s}})^T(\boldsymbol{q},\boldsymbol{p}) & (\partial_p \boldsymbol{P}_{\mathrm{p,s}})^T(\boldsymbol{q},\boldsymbol{p}) \end{pmatrix},$$

is symplectic, i.e., for any $(\boldsymbol{q}, \boldsymbol{p})$*,*

$$J_{\mathrm{p,s}}^T \begin{pmatrix} 0 & \mathrm{Id}_3 \\ -\mathrm{Id}_3 & 0 \end{pmatrix} J_{\mathrm{p,s}} = \begin{pmatrix} 0 & \mathrm{Id}_3 \\ -\mathrm{Id}_3 & 0 \end{pmatrix}, \tag{2.18}$$

where Id_3 *is a* 3×3 *identity matrix.*

Proposition 2.1.3 *The map* $\kappa_{\mathrm{p,s}}$ *is a canonical transform for any* $T, \delta > 0$*; furthermore it is bounded under sup norm.*

Proof: This is proposition 3.4 in [60] ■

For a canonical transform $\kappa_{\mathrm{p,s}}$ define the quantity $Z^{\kappa_{\mathrm{p,s}},t}(\mathbf{q}, \mathbf{p})$ for $|p| > 0$ as

$$Z^{\kappa_{\mathrm{p,s}},t} = \partial_{\mathbf{z}}(\mathbf{Q}(\mathbf{q}, \mathbf{p}) + \mathrm{i}\mathbf{P}(\mathbf{q}, \mathbf{p})) \tag{2.19}$$

With $\partial_{\mathbf{z}} = (\partial_{\mathbf{q}} - \mathrm{i}\partial_{\mathbf{p}})$. Dropping the superscript $\kappa_{\mathrm{p,s}}$, we

$$Z = (-\mathrm{iId}_3 \;\; \mathrm{Id}_3) \begin{pmatrix} \partial_{\mathbf{q}}\mathbf{Q} & \partial_{\mathbf{q}}\mathbf{P} \\ \partial_{\mathbf{p}}\mathbf{Q} & \partial_{\mathbf{p}}\mathbf{P} \end{pmatrix} \begin{pmatrix} \mathrm{iId}_3 \\ \mathrm{Id}_3 \end{pmatrix} \tag{2.20}$$

Definition 2.1.4 *The following notation will be useful. For* $\mathbf{u} \in C^\infty(S, \mathbb{C})$*, define for* $k \in \mathbb{N}$*.*

$$\Lambda_{k,S}(\mathbf{u}) = \max_{|\alpha_{\mathbf{p}}|+|\alpha_{\mathbf{q}}|=k} \sup_{(\mathbf{q},\mathbf{p})\in S} |\partial_{\mathbf{q}}^{\alpha_{\mathbf{q}}} \partial_{\mathbf{p}}^{\alpha_{\mathbf{p}}} \mathbf{u}(\mathbf{q}, \mathbf{p})| \tag{2.21}$$

with $\alpha_{\mathbf{q}}$*, and* $\alpha_{\mathbf{q}}$ *being multi-indices corresponding to* $\mathbf{q}$ *and* $\mathbf{p}$ *respectively.*

We will need the following lemma,

Lemma 2.1.5 *$Z^{\kappa_{\mathrm{p,s},t}}$ is invertible for $(\mathbf{q},\mathbf{p})\in\mathbb{R}^{2d}$ with $|\mathbf{p}|>0$. Furthermore, for any $k\geq 0$ and $\delta>0$, there exist constants $C_{k,\delta}$ such that*

$$\Lambda_{k,K_\delta}\Big(\big(Z^{\kappa_{\mathrm{p,s},t}}(\mathbf{q},\mathbf{p})\big)^{-1}\Big)\leq C_{k,\delta} \tag{2.22}$$

Proof: The proof uses the property of they symplectic transform to bound the eigenvalues of $Z^{\kappa_{\mathrm{p,s},t}}(Z^{\kappa_{\mathrm{p,s},t}})^*$, see Lemma 5.1 in [60] for details. ■

Definition 2.1.6 *For simplicity and brevity of notation; it is said $f\sim g$ if*

$$\int_{\mathbb{R}^{3d}} f(\mathbf{y})e^{\mathrm{i}/\epsilon\phi(t,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}=\int_{\mathbb{R}^{3d}} g(\mathbf{y})e^{\mathrm{i}/\epsilon\phi(t,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} \tag{2.23}$$

Lemma 2.1.7 *For any vector $\boldsymbol{a}(\mathbf{y},\mathbf{q},\mathbf{p})=(a_j)$ and matrix $M(\mathbf{y},\mathbf{q},\mathbf{p})=(M_{ij})$ in Schwartz class, one has the following integration by parts formula in the component-wise form, with $\partial_{\boldsymbol{z}}=(\partial_{z_1},\partial_{z_2},\partial_{z_3})$,*

$$\begin{aligned} a_j(x-Q)_j\sim&-\epsilon\partial_{z_m}\big(a_jZ^{-1}_{jm}\big),\\ (x-Q)_jM_{jl}(x-Q)_l\sim&\epsilon\partial_{z_m}Q_jM_{jl}Z^{-1}_{lm}+\mathcal{O}(\epsilon^2),\end{aligned} \tag{2.24}$$

Proof: This is a special case of Lemma 5.2 in [60] and it is Lemma 3.2 in [58] we refer to these for the detailed proof. ■

Theorem 2.1.8 *Given the Cauchy problem (2.2) in terms of the matrix system (2.5) with asymptotically high frequency initial condition X_0^ϵ, the following estimate holds*

$$\|X-X_{\mathrm{F},0}\|_{\mathrm{L}^2}\lesssim\epsilon\|X_0^\epsilon\|_{\mathrm{L}^2} \tag{2.25}$$

Proof: This is the main content of [59], X_{F} is defined as

$$X_{\mathrm{F},0} = (2\pi\epsilon)^{-3d/2} \sum_{n=1}^{7} \int_{\mathbb{R}^{3d}} \sigma_n(t,\mathbf{q},\mathbf{p}) R_n(t,\mathbf{Q}_n,\mathbf{P}_n) L_n(t,\mathbf{q},\mathbf{p})^T \times e^{\mathrm{i}/\epsilon \phi_n(t,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})} X_0^{\epsilon}(\mathbf{y})\, \mathrm{d}\mathbf{y}\, \mathrm{d}\mathbf{q}\, \mathrm{d}\mathbf{p} \quad (2.26)$$

sub-scripting over n (instead of $\mathrm{p}\pm, \mathrm{s}\pm, 0$) for the eigenfunctions and left and right eigenvectors defined from Eqs. (2.9)-(2.12) with $\mathbf{Q}_n = \mathbf{Q}_n(t,\mathbf{q},\mathbf{p})$, $\mathbf{P}_n = \mathbf{P}_n(t,\mathbf{q},\mathbf{p})$ and $\sigma_n(t,\mathbf{q},\mathbf{p})$ solving the evolution equation

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\sigma_n(t,\mathbf{q},\mathbf{p}) + \sigma_n(t,\mathbf{q},\mathbf{p})\lambda_n(t,\mathbf{q},\mathbf{p}) = 0 \\ \sigma_n(0,\mathbf{q},\mathbf{p}) = 2^{d/2} \end{cases} \quad (2.27)$$

where

$$\begin{aligned} \lambda_n = {} & L^T\big(\partial_{P_n} H_n \cdot \partial_{Q_n} R_n - \partial_{Q_n} H_n \cdot \partial_{P_n} R_n\big) \\ & - (\partial_{z_k} L_n)^T \Big(M_j - \partial_{P_{n,j}} H_n + \mathrm{i}(\partial_{Q_{n,j}} H_n - P_{n,l}\partial_{Q_{n,j}} M_l)\Big) R_n Z^{-1}_{n,jk} \\ & + \partial_{z_s} Q_{n,j} Z^{-1}_{n,ks} L^T \Big(- \partial_{Q_{m,j}} M_k + \frac{\mathrm{i}}{2} P_{n,l} \partial^2_{Q_{n,j},Q_{n,k}} M_l\Big) R_n \end{aligned} \quad (2.28)$$

With Q_n, P_n evaluated at $(t,\mathbf{q},\mathbf{p})$ and M_j evaluated at $\mathbf{Q}_n$ and H_n, L_n, R_n evaluated at $(\mathbf{Q}_n, \mathbf{P}_n)$. ■

It is noted that the work done in [59] for the first ordered FGA is the same as the operator is diagonalizable and the eigenspaces are non degenerate. X_{F} is defined here to show first order asymptotic equivalence; see corollary 1, to the FGA in section 2.3 using the amplitude factor derived from projections.

2.2 Decomposition and initial condition

The solution can be split into 2 waves each with 2 branches, so there are four branches. The FGA approximates the wavefield in Equation (1.4) by the following integral transform

$$\mathbf{u}_{\mathrm{F},0}(t,\mathbf{x}) = \frac{1}{(2\pi\epsilon)^{\frac{3d}{2}}}\int_{\mathbb{R}^{3d}} \sum_{\substack{w=\{p,s\}\\ b=\{\pm\}}} \mathbf{a}_{w,b}(t,\mathbf{y},\mathbf{p},\mathbf{q}) e^{\frac{\mathrm{i}}{\epsilon}\Phi_{w,b}(t,\mathbf{x},\mathbf{y},\mathbf{p},\mathbf{q})}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{q} \tag{2.29}$$

Where the superscript $w \in \{s,p\}$ represents the S-wave, P-wave dynamics respectively and $b \in \{+,-\}$ represents the respective $+$ or $-$ branch from the ray dynamics. The phase function is the same as eq. (1.39) execpt instead of suscrupted over n, the wave type and branch are subscripted; i.e.,

$$\Phi_{w,b}(t,\mathbf{x},\mathbf{y},\mathbf{p},\mathbf{q}) = \mathbf{P}_{w,b}\cdot(\mathbf{x}-\mathbf{Q}_{w,b}) + \frac{\mathrm{i}}{2}|\mathbf{x}-\mathbf{Q}_{w,b}|^2 - \mathbf{p}\cdot(\mathbf{y}-\mathbf{q}) + \frac{\mathrm{i}}{2}|\mathbf{y}-\mathbf{q}|^2 \tag{2.30}$$

with $\mathbf{P}_{w,b} = \mathbf{P}_{w,b}(t,\mathbf{q},\mathbf{p})$, $\mathbf{Q}^w_b = \mathbf{Q}_{w,b}(t,\mathbf{q},\mathbf{p})$. Note that if $t=0$

$$\Phi_{w,b}(0,\mathbf{x},\mathbf{y},\mathbf{p},\mathbf{q}) = \mathbf{p}\cdot(\mathbf{x}-\mathbf{q}) + \frac{\mathrm{i}}{2}|\mathbf{x}-\mathbf{q}|^2 - \mathbf{p}\cdot(\mathbf{y}-\mathbf{q}) + \frac{\mathrm{i}}{2}|\mathbf{y}-\mathbf{q}|^2. \tag{2.31}$$

is independent the branch and dynamic. Thus $\mathbf{u}(0,\mathbf{x}) = \mathbf{u}_{\mathrm{F}}(0,\mathbf{x})$ as the integral representation (2.29) is similar to the standard wave packet decomposition [25]. For the amplitude factor $\mathbf{a}_{w,b}(t,\mathbf{p},\mathbf{q})$ are given by

$$\mathbf{a}_{\mathrm{s},\pm}(t,\mathbf{y},\mathbf{q},\mathbf{p}) = a_{\mathrm{s},\pm}(t,\mathbf{q},\mathbf{p})\alpha_{\mathrm{s}}(\mathbf{y},\mathbf{q},\mathbf{p})\mathbf{P}_{\mathrm{s},\pm}(t,\mathbf{q},\mathbf{p}) \tag{2.32}$$

$$\mathbf{a}_{\mathrm{p},\pm}(t,\mathbf{y},\mathbf{q},\mathbf{p}) = a_{\mathrm{p},\pm}(t,\mathbf{q},\mathbf{p})\alpha_{\mathrm{p}}(\mathbf{y},\mathbf{q},\mathbf{p})\mathbf{P}_{\mathrm{p},\pm}(t,\mathbf{q},\mathbf{p}). \tag{2.33}$$

The equations for the prefactor $a_{w,\pm}$ are be derived by the asymptotic expansion. The amplitude for the S-waves need to be decomposed into $\mathbf{a}_{\mathrm{sh},\pm}$ and $\mathbf{a}_{\mathrm{sh},\pm}$, which represent the amplitude of the SV-waves and SH-waves respectively. $\mathbf{a}_{\mathrm{sh},\pm}$ and $\mathbf{a}_{\mathrm{sv},\pm}$ share the same dynamics and $\mathbf{a}_{\mathrm{s},\pm} = a_{\mathrm{sh},\pm}\alpha_{\mathrm{sh}}\hat{\mathbf{N}}_{\mathrm{sh}} + a_{\mathrm{sv},\pm}\alpha_{\mathrm{sv}}\hat{\mathbf{N}}_{\mathrm{sv}}$. The polarized directions of SV- and SH-waves are represented by $\hat{\mathbf{N}}_{\mathrm{sv}}$ and $\hat{\mathbf{N}}_{\mathrm{sh}}$ respectively. It is remarked that $\{\hat{\mathbf{N}}_{\mathrm{sh}}, \hat{\mathbf{N}}_{\mathrm{sv}}, \hat{\mathbf{P}}_{\mathrm{s}}\}$ form an orthonormal frame in $\mathbb{R}^3$. The prefactor amplitudes $\mathbf{a}_{\mathrm{p,s}}(t, \mathbf{q}, \mathbf{p})$ satisfy the following equations

$$\frac{\mathrm{d}a_{\mathrm{p},\pm}}{\mathrm{d}t} = a_{\mathrm{p},\pm}\left(\pm\frac{\partial_{\mathbf{Q}} c_{\mathrm{p}} \cdot \mathbf{P}_{\mathrm{p},\pm}}{|\mathbf{P}_{\mathrm{p},\pm}|} + \frac{1}{2}\,\mathrm{tr}\left(Z_{\mathrm{p},\pm}^{-1}\frac{\mathrm{d}Z_{\mathrm{p},\pm}}{\mathrm{d}t}\right)\right), \tag{2.34}$$

$$\frac{\mathrm{d}a_{\mathrm{sv},\pm}}{\mathrm{d}t} = a_{\mathrm{sv},\pm}\left(\pm\frac{\partial_{\mathbf{Q}} c_{\mathrm{s}} \cdot \mathbf{P}_{\mathrm{s},\pm}}{|\mathbf{P}_{\mathrm{s},\pm}|} + \frac{1}{2}\,\mathrm{tr}\left(Z_{\mathrm{s},\pm}^{-1}\frac{\mathrm{d}Z_{\mathrm{s},\pm}}{\mathrm{d}t}\right)\right) \mp a_{\mathrm{sh},\pm}\frac{\mathrm{d}\hat{\mathbf{N}}_{\mathrm{sh},\pm}}{\mathrm{d}t} \cdot \hat{\mathbf{N}}_{\mathrm{sv},\pm}, \tag{2.35}$$

$$\frac{\mathrm{d}a_{\mathrm{sh},\pm}}{\mathrm{d}t} = a_{\mathrm{sh},\pm}\left(\pm\frac{\partial_{\mathbf{Q}} c_{\mathrm{s},\pm} \cdot \mathbf{P}_{\mathrm{s},\pm}}{|\mathbf{P}_{\mathrm{s},\pm}|} + \frac{1}{2}\,\mathrm{tr}\left(Z_{\mathrm{s},\pm}^{-1}\frac{\mathrm{d}Z_{\mathrm{s},\pm}}{\mathrm{d}t}\right)\right) \pm a_{\mathrm{sv}}\frac{\mathrm{d}\hat{\mathbf{N}}_{\mathrm{sh},\pm}}{\mathrm{d}t} \cdot \hat{\mathbf{N}}_{\mathrm{sv},\pm}. \tag{2.36}$$

With short-hand notations from eqs. (2.19). Note that, the prefactor equation (2.34) is consistent with the one for acoustic wave equation with $c^2 = (\lambda + 2\mu)/\rho$ [11], and the last terms on the right-hand-side of (2.35)-(2.36) indicate the diabatic coupling of the polarized directions for SH- and SV-waves. This coupling is closely connected to the concept of Berry phase studied in quantum mechanics and topology (Chern number) [8, 82]. $Z_{w,\pm}$ is defined from (2.20). For notation, define the following quantities;

$$\psi_{w,\pm}(\mathbf{q}, \mathbf{p}) = \int_{\mathbb{R}^d} \alpha_{w,\pm} G_0^{\epsilon}(\mathbf{q}, \mathbf{p}, \mathbf{y})\ \mathrm{d}\mathbf{y} \tag{2.37}$$

$$G_0^{\epsilon}(\mathbf{q}, \mathbf{p}, \mathbf{y}) = \exp\left(\mathrm{i}/\epsilon(-\mathbf{p}\cdot(\mathbf{y}-\mathbf{q}) + \frac{\mathrm{i}}{2}|\mathbf{y}-\mathbf{q}|^2)\right) \tag{2.38}$$

$$G_{w,\pm}^{\epsilon}(t, \mathbf{q}, \mathbf{p}, \mathbf{x}, \mathbf{y}) = \exp\left(\mathrm{i}/\epsilon(\mathbf{P}_{w,\pm}\cdot(\mathbf{x}-\mathbf{Q}_{w,\pm}) + \frac{\mathrm{i}}{2}|\mathbf{x}-\mathbf{Q}_{w,\pm}|^2)\right) G_0^{\epsilon}(\mathbf{q}, \mathbf{p}, \mathbf{y}) \tag{2.39}$$

The factor of $\alpha_{w,\pm}(\mathbf{y},\mathbf{q},\mathbf{p})$ is determined by the initial conditions, which is

$$\alpha_{w,\pm}(\mathbf{y},\mathbf{q},\mathbf{p}) = \frac{2^{d/2-1}}{c_w|\mathbf{p}|^3}(\mathbf{u}_0^\epsilon c_w|\mathbf{p}| \pm \mathrm{i}\epsilon\mathbf{u}_1^\epsilon)\cdot\hat{\mathbf{n}}_w, \tag{2.40}$$

$G_0(\mathbf{q},\mathbf{p})$ is independent of the dynamics. With these definitions the FGA ansatz (2.41) can be written as

$$\mathbf{u}_{\mathrm{F},0}(t,\mathbf{x}) = \left(\frac{1}{2\pi\epsilon}\right)^{\frac{3d}{2}} \int_{\mathbb{R}^{3d}} \sum_{\substack{b=\pm \\ w=\mathrm{p,s}}} \mathbf{a}_{w,b} G^\epsilon_{w,\pm}(t,\mathbf{q},\mathbf{p},\mathbf{x},\mathbf{y})\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{q} \tag{2.41}$$

$$= \left(\frac{1}{2\pi\epsilon}\right)^{\frac{3d}{2}} \int_{\mathbb{R}^{2d}} \sum_{\substack{b=\pm \\ w=\mathrm{p,s}}} \mathbf{a}_{w,b}\psi_{w,\pm}(\mathbf{q},\mathbf{p}) \frac{G^\epsilon_{w,\pm}(t,\mathbf{q},\mathbf{p},\mathbf{x},\mathbf{y})}{G^\epsilon_0(\mathbf{q},\mathbf{p},\mathbf{y})}\,\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{q} \tag{2.42}$$

This decomposition is useful for implementation to incorporate the initial conditions and gives insight into the FGA integral representation itself. Dropping the super-script ϵ, denote the initial conditions can be written as,

$$\mathbf{u}(0,\mathbf{x}) = \mathbf{u}_0(\mathbf{x}) = \frac{2^{d/2}}{(2\pi\epsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} \mathbf{u}_0(\mathbf{y}) e^{\frac{\mathrm{i}}{\epsilon}\Phi}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} \tag{2.43}$$

$$\partial_t\mathbf{u}(0,\mathbf{x}) = \mathbf{u}_1(\mathbf{x}) = \frac{2^{d/2}}{(2\pi\epsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} \mathbf{u}_1(\mathbf{y}) e^{\frac{\mathrm{i}}{\epsilon}\Phi}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} \tag{2.44}$$

For the initial conditions it can be written,

$$2^{d/2}\mathbf{u}_0(\mathbf{y}) \sim \mathbf{a}_{\mathrm{p},+}(0,\mathbf{y},\mathbf{q},\mathbf{p}) + \mathbf{a}_{\mathrm{s},+}(0,\mathbf{y},\mathbf{q},\mathbf{p}) + \mathbf{a}_{\mathrm{s},+}(0,\mathbf{y},\mathbf{q},\mathbf{p}) + \mathbf{a}_{\mathrm{s},-}(0,\mathbf{y},\mathbf{q},\mathbf{p}) \tag{2.45}$$

Taking ∂_t,

$$2^{d/2}\mathbf{u}_1(\mathbf{y}) \sim \partial_t\big(\mathbf{a}_{\mathrm{p},+} + \mathbf{a}_{\mathrm{s},+} + \mathbf{a}_{\mathrm{p},-} + \mathbf{a}_{\mathrm{s},-}\big) + \frac{\mathrm{i}}{\epsilon}\partial_t\Phi\big(\mathbf{a}_{\mathrm{p},+} + \mathbf{a}_{\mathrm{s},+} + \mathbf{a}_{\mathrm{p},-} + \mathbf{a}_{\mathrm{s},-}\big) \tag{2.46}$$

Now

$$\partial_t \Phi_\pm(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p}) = \mp\Big(\partial c|\mathbf{p}| + \mathrm{i}c\frac{\mathbf{p}}{|\mathbf{p}|}\Big)\cdot(\mathbf{x}-\mathbf{q}) \mp c|\mathbf{p}| \tag{2.47}$$

and looking at the $\mathcal{O}(\epsilon^{-1})$ terms.

$$2^{d/2}\mathbf{u}_1(\mathbf{y}) \sim -\mathbf{a}_{\mathrm{p},+}c_\mathrm{p}|\mathbf{p}| - \mathbf{a}_{\mathrm{s},+}c_\mathrm{s}|\mathbf{p}| + \mathbf{a}_{\mathrm{p},-}c_\mathrm{p}|\mathbf{p}| + \mathbf{a}_{\mathrm{s},-}c_\mathrm{s}|\mathbf{p}| \tag{2.48}$$

Considering the integrand, dotting (2.45), (2.48) with $\mathbf{p}$ and multiplying $c_\mathrm{p}|\mathbf{p}|$ to (2.45), we arrive at

$$2^{d/2}\mathbf{u}_0 c_\mathrm{p}|\mathbf{p}|\cdot\mathbf{p} = \alpha_{\mathrm{p},+}c_\mathrm{p}|\mathrm{p}|^3 + \alpha_{\mathrm{p},-}c_\mathrm{p}|\mathbf{p}|^3 \tag{2.49}$$

$$2^{d/2}\mathbf{u}_1\cdot\mathbf{p} = -\alpha_{\mathrm{p},+}c_\mathrm{p}|\mathbf{p}|^3 + \alpha_{\mathrm{p},-}c_\mathrm{p}|\mathbf{p}|^3 \tag{2.50}$$

Solving for $\alpha_{\mathrm{p},\pm}$

$$\alpha_{\mathrm{p},\pm} = \frac{2^{d/2-1}(\mathbf{u}_0 c_\mathrm{p}|\mathbf{p}| \pm \mathrm{i}\epsilon\mathbf{u}_1)\cdot\mathbf{p}}{c_\mathrm{p}|\mathbf{p}|^3} \tag{2.51}$$

The equations for $\alpha_{\mathrm{sh},\pm}$ and $\alpha_{\mathrm{sv},\pm}$ are derived in the same fashion.

2.3 Derivation of the Evolution Equation

Definition 2.3.1 *The first order FGA is defined as*

$$\begin{aligned} u_{\mathrm{F},0}(t,\mathbf{x}) = (2\pi\epsilon)^{-3d/2}\int_{\mathbb{R}^{3d}}\sum_{b=\pm}\mathbf{a}_{\mathrm{p},b,0}(t,\mathbf{y},\mathbf{q},\mathbf{p})G^\epsilon_{\mathrm{p},b}(t,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p}) \\ + (\mathbf{a}_{\mathrm{sh},b,0}(t,\mathbf{y},\mathbf{q},\mathbf{p}) + \mathbf{a}_{\mathrm{sv},b,0}(t,\mathbf{y},\mathbf{q},\mathbf{p}))G^\epsilon_{\mathrm{s},b}(t,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{q} \end{aligned} \tag{2.52}$$

For the general computation we will drop the subscript and branch notation. The calculations for the two branches are identical for both P,S waves. Note that eq. (2.2) can be

rewritten as the following;

$$\rho(\boldsymbol{x})\partial_t^2\mathbf{u} = (\lambda+2\mu)\nabla(\nabla\cdot\mathbf{u}) - \mu\nabla\times\nabla\times\mathbf{u}. \tag{2.53}$$

From the coming asymototics, it is noted that if λ and μ are not constant they will not contribute to the first ordered FGA approximation. Both the isotropic and anisotropic cases have the same first ordered FGA approximation. This is not true for a higher ordered FGA approximation. Eq. (2.53) is linear, and thus one can derive the prefactor equations for P- and S-waves individually by assuming $\boldsymbol{A}_{\mathrm{p}} \parallel \boldsymbol{P}$ or $\boldsymbol{A}_{\mathrm{s}} \perp \boldsymbol{P}$, with the following Gaussian wave packet solution ansatz;

$$\mathbf{u}_{\mathrm{p,s}}(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p}) = \boldsymbol{A}_{\mathrm{p,s}}(t,\boldsymbol{q},\boldsymbol{p})\exp\big(\mathrm{i}/\epsilon\Phi_{\mathrm{p,s}}(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})\big). \tag{2.54}$$

For easier readability for further computations in this section (2.3), define $1/\epsilon = k$, a nondimensionalized wave number. Without loss of generality, we first consider the prefactor equation for the P-wave, with the governing equations for $\boldsymbol{Q}_{\mathrm{p}}$ and $\boldsymbol{P}_{\mathrm{p}}$ given by eq. (2.14). Plugging eq. (2.54) into eq. (2.53) and expanding the asymptotics in the weak sense of (2.23) yield

$$\begin{aligned}\rho\left(\boldsymbol{A}_{tt} + 2\mathrm{i}k\boldsymbol{A}_t\Phi_t + \mathrm{i}k\boldsymbol{A}\Phi_{tt} - k^2\boldsymbol{A}\Phi_t^2\right) \sim (\lambda+2\mu)\left(\mathrm{i}k\nabla(\boldsymbol{A}\cdot\nabla\Phi) - k^2(\boldsymbol{A}\cdot\nabla\Phi)\nabla\Phi\right)\\ -\mu\left(\mathrm{i}k\,\nabla\times(\nabla\Phi\times\boldsymbol{A}) - k^2\nabla\Phi\times(\nabla\Phi\times\boldsymbol{A})\right).\end{aligned} \tag{2.55}$$

The spatial and temporal derivatives of Φ are given by

$$\begin{aligned}
&\nabla\Phi = \mathbf{P} + \mathrm{i}(\boldsymbol{x} - \mathbf{Q}),\ \Delta\Phi = 3\mathrm{i},\ \nabla^2\Phi = \mathrm{i}\mathrm{Id}_3,\\
&|\nabla\Phi|^2 = |\mathbf{P}|^2 + 2\mathrm{i}\mathbf{P}\cdot(\boldsymbol{x} - \mathbf{Q}) - |\boldsymbol{x} - \mathbf{Q}|^2,\\
&\Phi_t = \ (\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\cdot(\boldsymbol{x} - \mathbf{Q}) - \mathbf{P}\cdot\mathbf{Q}_t,\\
&\Phi_t^2 = \ \big[(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\cdot(\boldsymbol{x} - \mathbf{Q})\big]^2 + (\mathbf{P}\cdot\mathbf{Q}_t)^2 - 2(\mathbf{P}\cdot\mathbf{Q}_t)\big[(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\cdot(\boldsymbol{x} - \mathbf{Q})\big],\\
&\Phi_{tt} = \ (\mathbf{P}_{tt} - \mathrm{i}\mathbf{Q}_{tt})\cdot(\boldsymbol{x} - \mathbf{Q}) - (\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\cdot\mathbf{Q}_t - \mathbf{P}_t\cdot\mathbf{Q}_t - \mathbf{P}\cdot\mathbf{Q}_{tt}.
\end{aligned} \tag{2.56}$$

Notice that the terms containing $k(\boldsymbol{x} - \boldsymbol{Q})$ will be of $\mathcal{O}(1)$ by the lemma of integration by parts, and for P-waves, $\boldsymbol{P}\times\boldsymbol{A} = 0$ and $\nabla\times\big((\boldsymbol{x} - \mathbf{Q})\times\boldsymbol{A}\big) = -2\boldsymbol{A}$. Plugging the derivatives of Φ in eq. (2.56) into eq. (2.55) produces, after neglecting the $\mathcal{O}(1)$ and lower order terms,

$$\begin{aligned}
2k\rho\boldsymbol{A}_t\big(\mathbf{P}\cdot\mathbf{Q}_t\big) \sim\ & k\rho\boldsymbol{A}\Big(- (\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\cdot\mathbf{Q}_t - \mathbf{P}_t\cdot\mathbf{Q}_t - \mathbf{P}\cdot\mathbf{Q}_{tt}\Big)\\
&+ \mathrm{i}k\rho\boldsymbol{A}\Big((\boldsymbol{x} - \mathbf{Q})\cdot\Big((\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\otimes(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\Big)(\boldsymbol{x} - \mathbf{Q})\Big)\\
&+ \mathrm{i}k^2\rho\boldsymbol{A}\Big(- 2(\mathbf{P}\cdot\mathbf{Q}_t)\big((\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\cdot(\boldsymbol{x} - \mathbf{Q})\big) + (\mathbf{P}\cdot\mathbf{Q}_t)^2\Big)\\
&+ (\lambda + 2\mu)\Big(- \mathrm{i}k\boldsymbol{A} + k^2\big((\boldsymbol{A}\cdot\mathbf{P})(\boldsymbol{x} - \mathbf{Q})\\
&\quad + (\boldsymbol{A}\otimes\mathbf{P})(\boldsymbol{x} - \mathbf{Q})\big) + \mathrm{i}\big((\boldsymbol{x} - \mathbf{Q})\otimes(\boldsymbol{x} - \mathbf{Q})\big)\boldsymbol{A}\Big)\\
&- 2\mathrm{i}\mu k\boldsymbol{A} - \mathrm{i}\mu k^2\Big(((\boldsymbol{x} - \mathbf{Q})\cdot\boldsymbol{A})(\boldsymbol{x} - \mathbf{Q}) - |\boldsymbol{x} - \mathbf{Q}|^2\boldsymbol{A}\Big)\\
&- \mu k^2\Big((\mathbf{P}\cdot\boldsymbol{A})(\boldsymbol{x} - \mathbf{Q}) - (\mathbf{P}\cdot(\boldsymbol{x} - \mathbf{Q}))\boldsymbol{A}\Big),
\end{aligned} \tag{2.57}$$

where $\otimes$ means the tensor product, e.g., $(\boldsymbol{A}\otimes\boldsymbol{P})_{jl} = A_jP_l$.

Expanding $\rho(\boldsymbol{x})$ around $\mathbf{Q}$ and truncating at order third order,

$$\rho(\boldsymbol{x}) = \rho + \partial_{\mathbf{Q}}\rho\cdot(\boldsymbol{x} - \mathbf{Q}) + \frac{1}{2}(\boldsymbol{x} - \mathbf{Q})\cdot\partial^2_{\mathbf{QQ}}\rho(\boldsymbol{x} - \mathbf{Q}), \tag{2.58}$$

and noticing that $\rho(\boldsymbol{Q})c^2(\boldsymbol{Q}) = \lambda + 2\mu$ is constant, one has

$$\partial_{\mathbf{Q}}\rho c^2 + 2c\partial_{\mathbf{Q}}c\rho = 0, \quad \partial_{\mathbf{Q}}\rho = -\frac{2c\partial_{\mathbf{Q}}c\rho}{c^2}. \tag{2.59}$$

Taking the second derivative for $\partial_{\mathbf{QQ}}\rho$ and substituting eq. (2.59) bring

$$\rho(\boldsymbol{x}) = \rho - \frac{2\partial_{\mathbf{Q}}c\rho}{c}\cdot(\boldsymbol{x}-\mathbf{Q}) + (\boldsymbol{x}-\mathbf{Q})\cdot\left(3\frac{\partial_{\mathbf{Q}}c\partial_{\mathbf{Q}}c^T\rho}{c^2} - \frac{\partial_{\mathbf{QQ}}c\rho}{c}\right)(\boldsymbol{x}-\mathbf{Q}). \tag{2.60}$$

Plugging eqs, (2.14) and (2.60) into eq. (2.57), and dividing by ρ yield, in componentwise form,

$$\begin{aligned}
2k\partial_t A_i c|\mathbf{P}| \sim & kA_i\Big(c(\partial_{\mathbf{Q}}c)_j P_j + \mathrm{i}c^2\Big) - \mathrm{i}c^2 kA_i \\
& + k^2 A_i(x-Q)_j M_{jl}(x-Q)_l - 2k^2 A_i c^2 P_j(x-Q)_j + c^2 N_i \\
& - \frac{2\mathrm{i}k\mu}{\rho}A_i - \frac{k^2\mu}{\rho}\Big(P_j A_j(x-Q)_i - P_j(x-Q)_j A_i\Big) \\
& - \frac{\mathrm{i}k^2\mu}{\rho}\Big((x-Q)_j A_j(x-Q)_i - |\boldsymbol{x}-\mathbf{Q}|^2 A_i\Big),
\end{aligned} \tag{2.61}$$

where M_{jl} and N_i are given as follows,

$$\begin{aligned}
M_{jl} = & -\mathrm{i}|\mathbf{P}|^2 c(\partial^2_{\mathbf{Q}}c)_{jk} + c\big(P_l + P_j\big)(\partial_{\mathbf{Q}}c)_j - \mathrm{i}c^2\frac{P_j P_l}{|\mathbf{P}|^2}, \\
N_i = & A_j P_j(x-Q)_i + A_i P_j(x-Q)_j + \mathrm{i}(x-Q)_i(x-Q)_j A_j.
\end{aligned} \tag{2.62}$$

Assuming that $\boldsymbol{A} = a_{\mathrm{p}}\hat{\boldsymbol{N}}_{\mathrm{p}}$, with $\hat{\boldsymbol{N}}_{\mathrm{p}} = \dfrac{\boldsymbol{P}}{|\boldsymbol{P}|}$, then

$$\partial_t\boldsymbol{A} = \partial_t a_{\mathrm{p}}\hat{\boldsymbol{N}}_{\mathrm{p}} + a_{\mathrm{p}}\partial_t\hat{\boldsymbol{N}}_{\mathrm{p}} = \partial_t a_{\mathrm{p}}\frac{\boldsymbol{P}}{|\boldsymbol{P}|} + a_{\mathrm{p}}\partial_t\Big(\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\Big). \tag{2.63}$$

Plugging this into eq. (2.61) with the ray equations (2.14), using the P-wave velocity

eq. (2.1) and grouping in powers of $(\boldsymbol{x}-\mathbf{Q})$ produce

$$\begin{aligned}
2k\partial_t a_{\mathrm{p}} c|\boldsymbol{P}|P_i \sim\ & -2ka_{\mathrm{p}}c|\mathbf{P}|^2\partial_t\Big(\frac{P_i}{|\boldsymbol{P}|}\Big) + ka_{\mathrm{p}}c(\partial_{\mathbf{Q}}c)_j P_j P_i \\
& - \frac{2\mathrm{i}k\mu P_i}{\rho}a_{\mathrm{p}} + k^2 a_{\mathrm{p}}|\mathbf{P}|^2\left(c^2-\frac{\mu}{\rho}\right)(x-Q)_i \\
& + k^2 a_{\mathrm{p}}\left(\frac{\mu}{\rho}-c^2\right)P_jP_i(x-Q)_j + k^2\frac{\mathrm{i}\mu a_{\mathrm{p}}}{\rho}|x-Q|^2P_i \\
& + k^2\mathrm{i}a_{\mathrm{p}}P_j\left(c^2-\frac{\mu}{\rho}\right)(x-Q)_j(x-Q)_i + k^2a_{\mathrm{p}}P_i(x-Q)_jM_{jl}(x-Q)_l.
\end{aligned}$$

Denoting $\partial_{\boldsymbol{z}} = (\partial_1,\partial_2,\partial_3)$ for an ease of notation, applying eq. (2.1.5) and considering only the $\mathcal{O}(k)$ terms;

$$\begin{aligned}
2\partial_t a_{\mathrm{p}} c|\boldsymbol{P}|P_i \sim\ & -2a_{\mathrm{p}}c|\mathbf{P}|^2\partial_t\Big(\frac{P_i}{|\boldsymbol{P}|}\Big) + a_{\mathrm{p}}c(\partial_{\mathbf{Q}}c)_jP_jP_i - \frac{2\mathrm{i}\mu a_{\mathrm{p}}}{\rho}P_i \\
& - \partial_l\left(\alpha\Big(|\mathbf{P}|^2-P_jP_i\Big)a_n\left(\frac{\mu}{\rho}-c^2\right)Z^{-1}_{jl}\right) \\
& + \mathrm{i}a_{\mathrm{p}}P_j\left(c^2-\frac{\mu}{\rho}\right)Z^{-1}_{il}\partial_l Q_j + \frac{\mathrm{i}\mu a_{\mathrm{p}}}{\rho}P_iZ^{-1}_{jl}\partial_lQ_j \\
& + a_{\mathrm{p}}P_i\left(c(\partial_{\mathbf{Q}}c)_jP_r + cP_j(\partial_{\mathbf{Q}}c)_j - \mathrm{i}c^2\frac{P_jP_r}{|\mathbf{P}|^2}\right)cZ^{-1}_{rl}\partial_lQ_j \\
& - a_{\mathrm{p}}\mathrm{i}P_i|\mathbf{P}|^2c(\partial^2_{\mathbf{Q}}c)_{jr}Z^{-1}_{rl}\partial_lQ_j.
\end{aligned} \tag{2.64}$$

To derive an ordinary differential equation (ODE) instead of a partial differential equation (PDE) for a_{p}, one needs to simplify the terms containing $\partial_k a_{\mathrm{p}}$ as

$$\Big(|\mathbf{P}|^2-P_jP_i\Big)\partial_l\left(a_{\mathrm{p}}\left(\frac{\mu}{\rho}-c^2\right)Z^{-1}_{jl}\right). \tag{2.65}$$

Recall that eq. (2.64) holds in the sense of integral form (2.23), and now we shall consider a strong form of eq. (2.64), i.e., equate the integrands of the integrals on both sides. After

taking the dot product of integrands with $\mathbf{P}$, one has

$$
\begin{aligned}
2\frac{\partial_t a_{\mathrm{p}}}{a_{\mathrm{p}}} = {} & 2\frac{(\partial_{\mathbf{Q}} c)_j P_j}{|\mathbf{P}|} - \frac{2\mu \mathrm{i}}{\rho c|\mathbf{P}|} + \frac{\mathrm{i}\mu}{\rho c|\mathbf{P}|}\left(\delta_{ij} - \frac{P_j P_i}{|\mathbf{P}|^2}\right) Z_{il}^{-1}\partial_l Q_j \\
& + \frac{1}{c}\left(c^2 - \frac{\mu}{\rho}\right)\frac{P_i}{|\mathbf{P}|}\partial_l\left(\frac{P_j P_i}{|\mathbf{P}|^2}\right) Z_{jl}^{-1} + \frac{1}{|\mathbf{P}|}(\partial_{\mathbf{Q}} c)_j P_k Z_{kl}^{-1}\partial_l Q_j \\
& + \frac{1}{|\mathbf{P}|} P_j (\partial_{\mathbf{Q}} c)_j Z_{kl}^{-1}\partial_l Q_j - \mathrm{i}|\mathbf{P}|(\partial_{\mathbf{Q}}^2 c)_{jk} Z_{kl}^{-1}\partial_l Q_j,
\end{aligned}
$$

where the terms (2.65) actually become zero since $\mathbf{P}\cdot(|\mathbf{P}|^2 - \mathbf{P}\otimes\mathbf{P}) = 0$, and we have used the fact that

$$
\frac{P_i}{|\boldsymbol{P}|}\partial_t\Big(\frac{P_i}{|\boldsymbol{P}|}\Big) = \frac{1}{2}\Big(\frac{P_i}{|\boldsymbol{P}|}\partial_t\Big(\frac{P_i}{|\boldsymbol{P}|}\Big) + \partial_t\Big(\frac{P_i}{|\boldsymbol{P}|}\Big)\frac{P_i}{|\boldsymbol{P}|}\Big) = \partial_t\Big(\frac{P_i}{|\boldsymbol{P}|}\frac{P_i}{|\boldsymbol{P}|}\Big) = 0,
$$

which implies $\partial_t\Big(\dfrac{P_i}{|\boldsymbol{P}|}\Big)P_i = 0$.

Since $Z = \partial_{\mathbf{z}}(\mathbf{Q} + \mathrm{i}\mathbf{P})$ by eq. (2.19), $\partial_t Z = \partial_t\partial_{\mathbf{z}}\mathbf{Q} + \mathrm{i}\partial_t\partial_{\mathbf{z}}\mathbf{P}$. Then eq. (2.14) implies

$$
\begin{aligned}
\partial_t\partial_{\mathbf{z}}\mathbf{Q} &= \partial_{\mathbf{z}}\mathbf{Q}\frac{\partial_{\mathbf{Q}} c\otimes\mathbf{P}}{|\mathbf{P}|} + c\partial_{\mathbf{z}}\mathbf{P}\left(\frac{\mathrm{Id}_3}{|\mathbf{P}|} - \frac{\mathbf{P}\otimes\mathbf{P}}{|\mathbf{P}|^3}\right), \\
\partial_t\partial_{\mathbf{z}}\mathbf{P} &= -|\mathbf{P}|\partial_{\mathbf{z}}\mathbf{Q}\partial_{\mathbf{Q}}^2 c - \partial_{\mathbf{z}}\mathbf{P}\frac{\mathbf{P}\otimes\partial_{\mathbf{Q}} c}{|\mathbf{P}|}.
\end{aligned}
\tag{2.66}
$$

Using eq. (2.66) for further simplifications give

$$
\begin{aligned}
2\frac{\partial_t a_{\mathrm{p}}}{a_{\mathrm{p}}} = {} & 2\frac{(\partial_{\mathbf{Q}} c)_i P_i}{|\mathbf{P}|} + \mathrm{tr}\left(Z^{-1}\partial_t Z\right) \\
& + \frac{1}{c}\left(c^2 - \frac{\mu}{\rho}\right)\frac{P_i}{|\mathbf{P}|}\partial_l\left(\frac{P_j P_i}{|\mathbf{P}|^2}\right) Z_{jl}^{-1} - \frac{1}{c|\mathbf{P}|}\left(c^2 - \frac{\mu}{\rho}\right)\left(\delta_{ij} - \frac{P_j P_i}{|\mathbf{P}|^2}\right) Z_{il}^{-1}\partial_l P_j,
\end{aligned}
\tag{2.67}
$$

where the last two terms can be grouped as

$$
\frac{1}{c|\mathbf{P}|}\left(c^2 - \frac{\mu}{\rho}\right)\partial_l\left(P_i\Big(\frac{P_j P_i}{|\mathbf{P}|^2} - \delta_{ij}\Big)\right) Z_{jl}^{-1} = 0, \tag{2.68}
$$

which implies a nice ODE for a_{p} as in eq. (2.34),

$$\frac{\mathrm{d}a_{\mathrm{p}}}{\mathrm{d}t} = a_{\mathrm{p}}\left(\frac{\partial_{\mathbf{Q}} c\cdot\mathbf{P}}{|\mathbf{P}|} + \frac{1}{2}\,\mathrm{tr}\left(Z^{-1}\frac{\mathrm{d}Z}{\mathrm{d}t}\right)\right). \tag{2.69}$$

For the $\mathcal{O}(1)$ equation, from the expansion and similar steps we arrive at, in component form;

$$\begin{aligned}
2c|P|\frac{\mathrm{d}a_1}{\mathrm{d}t} &= 2a_1 c\partial c_j P_j|\mathbf{P}| + a_1 c|P|\big(Z^{-1}_{jk}\frac{\mathrm{d}Z_{jk}}{\mathrm{d}t}\big) && (2.70)\\
&- \imath(a_0)_{tt} - \imath a_0 - \partial_k\left(4\Big(a_0\frac{P_i}{|P|}\Big)_t P_i\partial c_j\mathcal{Z}^{-1}_{jk}\right)\frac{P_i}{|P|} && (2.71)\\
&- 2(a_0)_t|P|\left(3\frac{\partial c_j\partial c_k}{c} - \partial^2 c_{jk}\right)\mathcal{Z}^{-1}_{jl}\partial_l(Q_k)\\
&+ \partial_k\left(a_0\frac{P_i}{|P|}M_{jk}\mathcal{Z}^{-1}_{jk}\right)\frac{P_i}{|P|} + a_0 N_{jk}\mathcal{Z}^{-1}_{jl}\partial_l(Q_k)\\
&+ 2\imath\partial_n\left(a_0\frac{P_i}{|P|}\Big(\frac{\partial c_l\partial c_j\partial c_k|P|^2}{c} + 2\imath\partial c_l\partial c_k P_j - c\partial c_l\frac{P_jP_k}{|P|^2}\Big)\mathcal{Z}^{-1}_{jm}\partial_m(Q_k)\mathcal{Z}^{-1}_{ln}\right)\frac{P_i}{|P|}
\end{aligned}$$

Where

$$M_{jk} = 2\partial c_j|P| + 2\imath c\frac{P_j}{|P|} + \partial c_j\partial c_k P_k + c\partial^2 c_{kj}P_k + \imath\frac{2c\partial c_k P_kP_j}{|P|^2} + \imath c\partial c_j \tag{2.72}$$

$$\begin{aligned}
N_{jk} = 4\frac{\partial c_j\partial c_k}{c}|P| + 4\imath\frac{\partial c_k P_j}{|P|} - \partial^2 c_{jk}\partial c_n P_n - \imath c\partial^2 c_{jk}\\
+ \frac{\partial c_j\partial c_k\partial c_n P_n}{c} + 2\partial c_j\partial^2 c_{nk}P_n + 4\imath\frac{\partial c_j\partial c_n P_nP_k}{|P|^2} + \imath\partial c_j\partial c_k
\end{aligned} \tag{2.73}$$

We write this as

$$\frac{\mathrm{d}a_{\mathrm{p},1}}{\mathrm{d}t} = a_{\mathrm{p},1}\left(\frac{\partial_{\mathbf{Q}} c\cdot\mathbf{P}}{|\mathbf{P}|} + \frac{1}{2}\,\mathrm{tr}\left(Z^{-1}\frac{\mathrm{d}Z}{\mathrm{d}t}\right)\right) + F_p(a_{\mathrm{p},0},\mathbf{Q},\mathbf{P},c_{\mathrm{p}}). \tag{2.74}$$

with F_p containing first and second derivatives of its arguments, which are smooth for $|\mathbf{P}| > 0$.

Similarly, one can derive the prefactor equations for SV- and SH-waves by assuming $\boldsymbol{A} = a_{\mathrm{sv}}\hat{\boldsymbol{N}}_{\mathrm{sv}} + a_{\mathrm{sh}}\hat{\boldsymbol{N}}_{\mathrm{sh}}$ with $\hat{\boldsymbol{N}}_{\mathrm{sv}} \perp \boldsymbol{P}$, $\hat{\boldsymbol{N}}_{\mathrm{sh}} \perp \boldsymbol{P}$ and $\hat{\boldsymbol{N}}_{\mathrm{sv}} \perp \hat{\boldsymbol{N}}_{\mathrm{sh}}$ in eq. (2.54). The calculations will be essentially the same as the prefactor equation for P-waves except that one will have the diabatic coupling terms of $\hat{\boldsymbol{N}}_{\mathrm{sv}}$ and $\hat{\boldsymbol{N}}_{\mathrm{sh}}$ as shown below,

$$\begin{aligned}
\frac{\mathrm{d}a_{\mathrm{sv}}}{\mathrm{d}t} &= a_{\mathrm{sv}}\left(\frac{\partial_{\boldsymbol{Q}_{\mathrm{s}}} c_{\mathrm{s}} \cdot \boldsymbol{P}_{\mathrm{s}}}{|\boldsymbol{P}_{\mathrm{s}}|} + \frac{1}{2}\,\mathrm{tr}\left(Z_{\mathrm{s}}^{-1}\frac{\mathrm{d}Z_{\mathrm{s}}}{\mathrm{d}t}\right)\right) - a_{\mathrm{sh}}\left(\frac{\mathrm{d}\hat{\boldsymbol{N}}_{\mathrm{sh}}}{\mathrm{d}t}\cdot\hat{\boldsymbol{N}}_{\mathrm{sv}} + m_{\mathrm{sh}\to\mathrm{sv}}\right),\\
\frac{\mathrm{d}a_{\mathrm{sh}}}{\mathrm{d}t} &= a_{\mathrm{sh}}\left(\frac{\partial_{\boldsymbol{Q}_{\mathrm{s}}} c_{\mathrm{s}} \cdot \boldsymbol{P}_{\mathrm{s}}}{|\boldsymbol{P}_{\mathrm{s}}|} + \frac{1}{2}\,\mathrm{tr}\left(Z_{\mathrm{s}}^{-1}\frac{\mathrm{d}Z_{\mathrm{s}}}{\mathrm{d}t}\right)\right) - a_{\mathrm{sv}}\left(\frac{\mathrm{d}\hat{\boldsymbol{N}}_{\mathrm{sv}}}{\mathrm{d}t}\cdot\hat{\boldsymbol{N}}_{\mathrm{sh}} + m_{\mathrm{sv}\to\mathrm{sh}}\right),
\end{aligned} \tag{2.75}$$

where the interaction terms are given by

$$\begin{aligned}
m_{\mathrm{sh}\to\mathrm{sv}} &= \mathrm{i}\frac{\lambda+\mu}{\rho c_{\mathrm{s}}|P_{\mathrm{s}}|}\left(\hat{\boldsymbol{N}}_{\mathrm{sv}}\cdot(Z_{\mathrm{s}}^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}_s)\hat{\boldsymbol{N}}_{\mathrm{sh}} - \hat{\boldsymbol{N}}_{\mathrm{sh}}\cdot(Z_{\mathrm{s}}^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}_s)\hat{\boldsymbol{N}}_{\mathrm{sv}}\right),\\
m_{\mathrm{sv}\to\mathrm{sh}} &= -m_{\mathrm{sh}\to\mathrm{sv}}.
\end{aligned}$$

Also, note that by $\hat{\boldsymbol{N}}_{\mathrm{sv}} \perp \hat{\boldsymbol{N}}_{\mathrm{sh}}$, one has that $\dfrac{\mathrm{d}\hat{\boldsymbol{N}}_{\mathrm{sh}}}{\mathrm{d}t}\cdot\hat{\boldsymbol{N}}_{\mathrm{sv}} + \dfrac{\mathrm{d}\hat{\boldsymbol{N}}_{\mathrm{sv}}}{\mathrm{d}t}\cdot\hat{\boldsymbol{N}}_{\mathrm{sh}} = 0$.

Next, we shall show that $m_{\mathrm{sh}\to\mathrm{sv}} = m_{\mathrm{sv}\to\mathrm{sh}} = 0$ by proving that $Z_{\mathrm{s}}^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}_s$ is symmetric using the following argument. Eq. (2.18) implies, with the subscript s omitted for convenience,

$$\partial_{\boldsymbol{q}}\boldsymbol{Q}(\partial_{\boldsymbol{q}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{q}}\boldsymbol{P}(\partial_{\boldsymbol{q}}\boldsymbol{Q})^{\mathrm{T}} = 0_{3\times3}, \tag{2.76}$$

$$\partial_{\boldsymbol{q}}\boldsymbol{Q}(\partial_{\boldsymbol{p}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{q}}\boldsymbol{P}(\partial_{\boldsymbol{p}}\boldsymbol{Q})^{\mathrm{T}} = \mathrm{Id}_3, \tag{2.77}$$

$$\partial_{\boldsymbol{p}}\boldsymbol{Q}(\partial_{\boldsymbol{q}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{p}}\boldsymbol{P}(\partial_{\boldsymbol{q}}\boldsymbol{Q})^{\mathrm{T}} = -\mathrm{Id}_3, \tag{2.78}$$

$$\partial_{\boldsymbol{p}}\boldsymbol{Q}(\partial_{\boldsymbol{p}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{p}}\boldsymbol{P}(\partial_{\boldsymbol{p}}\boldsymbol{Q})^{\mathrm{T}} = 0_{3\times3}, \tag{2.79}$$

where $0_{3\times3}$ is 3-by-3 zero matrix.

Eq. (2.76)$-$i$\times$ Eq. (2.78) gives

$$\partial_{\boldsymbol{z}}\boldsymbol{Q}(\partial_{\boldsymbol{q}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{z}}\boldsymbol{P}(\partial_{\boldsymbol{q}}\boldsymbol{Q})^{\mathrm{T}} = \mathrm{i}\mathrm{Id}_3. \tag{2.80}$$

Eq. (2.77)$-$i$\times$ Eq. (2.79) gives

$$\partial_{\boldsymbol{z}}\boldsymbol{Q}(\partial_{\boldsymbol{p}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{z}}\boldsymbol{P}(\partial_{\boldsymbol{p}}\boldsymbol{Q})^{\mathrm{T}} = \mathrm{Id}_3. \tag{2.81}$$

Eq. (2.80)$-$i$\times$ Eq. (2.81) gives

$$\partial_{\boldsymbol{z}}\boldsymbol{Q}(\partial_{\boldsymbol{z}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{z}}\boldsymbol{P}(\partial_{\boldsymbol{z}}\boldsymbol{Q})^{\mathrm{T}} = 0_{3\times 3}.$$

Combined with $\partial_{\boldsymbol{z}}\boldsymbol{Q}(\partial_{\boldsymbol{z}}\boldsymbol{Q})^{\mathrm{T}} - \partial_{\boldsymbol{z}}\boldsymbol{Q}(\partial_{\boldsymbol{z}}\boldsymbol{Q})^{\mathrm{T}} = 0_{3\times 3}$, one has

$$\partial_{\boldsymbol{z}}QZ^{\mathrm{T}} - Z(\partial_{\boldsymbol{z}}Q)^{\mathrm{T}} = 0,$$

which implies $Z^{-1}\partial_{\boldsymbol{z}}Q = (\partial_{\boldsymbol{z}}Q)^{\mathrm{T}}(Z^{\mathrm{T}})^{-1} = (\partial_{\boldsymbol{z}}Q)^{\mathrm{T}}(Z^{-1})^{\mathrm{T}} = (Z^{-1}\partial_{\boldsymbol{z}}Q)^{\mathrm{T}}$. Therefore, $Z^{-1}\partial_{\boldsymbol{z}}Q$ is symmetric, and then $m_{\mathrm{sh}\to\mathrm{sv}} = m_{\mathrm{sv}\to\mathrm{sh}} = 0$, which brings eqs. (2.35) and (2.36) by eq. (2.75).

2.4 Proof of Convergence

In this section the asymptotic is expanded and done without the dynamics. The necessary operators are derived and proof of convergence is shown. Starting from equa-

tion (2.57)

$$\begin{aligned}
&\rho\Big(\boldsymbol{A}_{tt}+\frac{2}{\epsilon}\mathrm{i}\boldsymbol{A}_t\left((\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)\cdot(\boldsymbol{x}-\mathbf{Q})-\mathbf{P}\cdot\mathbf{Q}_t\right)\\
&\qquad\qquad+\frac{\mathrm{i}}{\epsilon}\boldsymbol{A}\left((\mathbf{P}_{tt}-\mathrm{i}\mathbf{Q}_{tt})\cdot(\boldsymbol{x}-\mathbf{Q})-(\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)\cdot\mathbf{Q}_t\right)\\
&-\frac{1}{\epsilon^2}\boldsymbol{A}\left(\left[(\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)\cdot(\boldsymbol{x}-\mathbf{Q})\right]^2+(\mathbf{P}\cdot\mathbf{Q}_t)^2-2(\mathbf{P}\cdot\mathbf{Q}_t)\left[(\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)\cdot(\boldsymbol{x}-\mathbf{Q})\right]\right)\Big)\\
&\sim(\lambda+2\mu)\left(\frac{\mathrm{i}}{\epsilon}\nabla(\boldsymbol{A}\cdot(\mathbf{P}+\mathrm{i}(\boldsymbol{x}-\mathbf{Q})))-\frac{1}{\epsilon^2}(\boldsymbol{A}\cdot(\mathbf{P}+\mathrm{i}(\boldsymbol{x}-\mathbf{Q})))(\mathbf{P}+\mathrm{i}(\boldsymbol{x}-\mathbf{Q}))\right)\\
&-\mu\left(\frac{\mathrm{i}}{\epsilon}\nabla\times((\mathbf{P}+\mathrm{i}(\boldsymbol{x}-\mathbf{Q}))\times\boldsymbol{A})-\frac{1}{\epsilon^2}\nabla(\mathbf{P}+\mathrm{i}(\boldsymbol{x}-\mathbf{Q}))\times((\mathbf{P}+\mathrm{i}(\boldsymbol{x}-\mathbf{Q}))\times\boldsymbol{A})\right).
\end{aligned}\tag{2.82}$$

Expanding $\rho(\boldsymbol{x})$ around $\mathbf{Q}$ and truncating at order third order, Grouping in terms of $(\mathbf{x}-\mathbf{Q})$, up to $\mathcal{O}(3)$, and dropping terms that produce terms higher than $\mathcal{O}(\epsilon)$, we can rewrite with simplifying, starting with $\mathcal{O}((\mathbf{x}-\mathbf{Q})^0)$

$$\rho M_0+\frac{\lambda}{\epsilon}\boldsymbol{A}+(\lambda+\mu)\frac{1}{\epsilon^2}(\boldsymbol{A}\cdot\mathbf{P})\mathbf{P}+\frac{\mu}{\epsilon^2}(\mathbf{P}\cdot\mathbf{P})\boldsymbol{A}\tag{2.83}$$

The $\mathcal{O}((\mathbf{x}-\mathbf{Q})^1)$ term is

$$\begin{aligned}
&\rho M_1\cdot(\boldsymbol{x}-\mathbf{Q})+(\lambda+\mu)\frac{\mathrm{i}}{\epsilon^2}(\boldsymbol{A}\cdot(\boldsymbol{x}-\mathbf{Q}))\mathbf{P}+(\lambda+\mu)\frac{\mathrm{i}}{\epsilon^2}(\boldsymbol{A}\cdot\mathbf{P})(\boldsymbol{x}-\mathbf{Q})\\
&\qquad\qquad+\frac{2\mu\mathrm{i}}{\epsilon^2}(\mathbf{P}\cdot(\boldsymbol{x}-\mathbf{Q}))\boldsymbol{A}+\partial_{\mathbf{Q}}\rho\cdot(\boldsymbol{x}-\mathbf{Q})M_0
\end{aligned}\tag{2.84}$$

The $\mathcal{O}((\mathbf{x}-\mathbf{Q})^2)$ term,

$$\rho\left(-\frac{1}{\epsilon^2}\boldsymbol{A}\big[(\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)\cdot(\boldsymbol{x}-\mathbf{Q})\big]^2\right) \\ -(\lambda+\mu)\frac{1}{\epsilon^2}(\boldsymbol{A}\cdot(\boldsymbol{x}-\mathbf{Q}))(\boldsymbol{x}-\mathbf{Q})-\frac{\mu}{\epsilon^2}|\boldsymbol{x}-\mathbf{Q}|^2\boldsymbol{A} \\ +\partial_{\mathbf{Q}}\rho\cdot(\boldsymbol{x}-\mathbf{Q})M_1\cdot(\boldsymbol{x}-\mathbf{Q})+\frac{1}{2}(\boldsymbol{x}-\mathbf{Q})\cdot\partial^2_{\mathbf{QQ}}\rho(\boldsymbol{x}-\mathbf{Q})M_0 \quad (2.85)$$

Finally the $\mathcal{O}((\mathbf{x}-\mathbf{Q})^3)$ term,

$$(\boldsymbol{x}-\mathbf{Q})\cdot\partial^2_{\mathbf{QQ}}\rho(\boldsymbol{x}-\mathbf{Q})\left(\frac{\boldsymbol{A}}{\epsilon^2}(\mathbf{P}\cdot\mathbf{Q}_t)\big[(\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)\cdot(\boldsymbol{x}-\mathbf{Q})\big]\right) \quad (2.86)$$

With M_0 and M_1 defined as

$$M_0=\boldsymbol{A}_{tt}-\boldsymbol{A}_t\frac{2\mathrm{i}}{\epsilon}\mathbf{P}\cdot\mathbf{Q}_t-\boldsymbol{A}\frac{\mathrm{i}}{\epsilon}(\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)\cdot\mathbf{Q}_t-\frac{1}{\epsilon^2}\boldsymbol{A}(\mathbf{P}\cdot\mathbf{Q}_t)^2 \\ M_1=\frac{2}{\epsilon}\mathrm{i}\boldsymbol{A}_t(\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)+\frac{\mathrm{i}}{\epsilon}\boldsymbol{A}(\mathbf{P}_{tt}-\mathrm{i}\mathbf{Q}_{tt})+\frac{2}{\epsilon}\boldsymbol{A}(\mathbf{P}\cdot\mathbf{Q}_t)\big[(\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)\big] \quad (2.87)$$

Now applying lemma (2.1.7), first applying integration by parts to $M_1\cdot(\mathbf{x}-\mathbf{Q})$ to clarify notation

$$M_1\cdot(\mathbf{x}-\mathbf{Q})\sim-\epsilon\partial_l(M_{1,ij}Z^{-1}_{jl})\sim-\epsilon\partial_{\mathbf{z}}:(Z^{-\mathrm{T}}M_1) \quad (2.88)$$

where : denotes a contraction of the indicates. Considering each order separately, first the $\mathcal{O}((\mathbf{x}-\mathbf{Q})^1)$

$$-\epsilon\rho\partial_{\mathbf{z}}:(\rho Z^{-\mathrm{T}}M_1)-(\lambda+\mu)\frac{\mathrm{i}}{\epsilon}\partial_{\mathbf{z}}:(Z^{-\mathrm{T}}\boldsymbol{A}\mathbf{P})-(\lambda+\mu)\frac{\mathrm{i}}{\epsilon}\partial_{\mathbf{z}}:(\boldsymbol{A}\cdot\mathbf{P}Z^{-\mathrm{T}}) \\ -\frac{2\mu\mathrm{i}}{\epsilon}\partial_{\mathbf{z}}:(Z^{-\mathrm{T}}\mathbf{P}\boldsymbol{A})-\epsilon\partial_{\mathbf{z}}:(Z^{-\mathrm{T}}\partial_{\mathbf{Q}}\rho M_0) \quad (2.89)$$

The $\mathcal{O}((\mathbf{x}-\mathbf{Q})^2)$ term

$$\begin{aligned}
&\frac{-\rho}{\epsilon}\boldsymbol{A}\,\mathrm{tr}\left((\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)\otimes(\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)Z^{-1}\partial_{\mathbf{z}}\mathbf{Q}\right)\\
&\qquad+\partial_{\mathbf{z}}:(\partial_{\mathbf{z}}:(\rho\boldsymbol{A}(\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)\otimes(\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)Z^{-1})Z^{-1})\\
&\quad-(\lambda+\mu)\frac{1}{\epsilon}Z^{-1}:(\partial_{\mathbf{z}}\mathbf{Q}\boldsymbol{A})-(\lambda+\mu)\partial_{\mathbf{z}}:(\partial_{\mathbf{z}}:(\partial_{\mathbf{z}}(\mathbf{Q}A)Z^{-1})Z^{-1})\\
&\qquad-\frac{\mu}{\epsilon}\boldsymbol{A}\,\mathrm{tr}(Z^{-1}\partial_{\mathbf{z}})-\mu\partial_{\mathbf{z}}:(\partial_{\mathbf{z}}:(Z^{-1})\boldsymbol{A}Z^{-1})\\
&\qquad+\epsilon\partial_{\mathbf{Q}}\rho M_1\,\mathrm{tr}(Z^{-1}\partial_{\mathbf{z}})+\epsilon\frac{1}{2}\partial^2_{\mathbf{QQ}}\rho M_0\,\mathrm{tr}(Z^{-1}\partial_{\mathbf{z}})\\
&\qquad+\epsilon^2\partial_{\mathbf{z}}:(\partial_{\mathbf{z}}:(Z^{-1}\partial^2_{\mathbf{QQ}}\rho)M_0Z^{-1})
\end{aligned}\tag{2.90}$$

The $\mathcal{O}((\mathbf{x}-\mathbf{Q})^3)$ term only up to first order terms

$$\partial_{\mathbf{z}}:\left(\mathrm{tr}(\partial^2_{\mathbf{QQ}}\rho Z^{-t}\partial_{\mathbf{z}}\mathbf{Q})(\mathbf{P}\cdot\mathbf{Q}_t)\left[(\mathbf{P}_t-\mathrm{i}\mathbf{Q}_t)\right]\boldsymbol{A}\right)\tag{2.91}$$

Expanding M_0 and M_1 when needed and grouping in terms of ϵ

$$
\begin{aligned}
&\frac{1}{\epsilon^2}\Big[-\rho\boldsymbol{A}(\mathbf{P}\cdot\mathbf{Q}_t)^2 + (\lambda+\mu)(\boldsymbol{A}\cdot\mathbf{P})\mathbf{P} + \mu(\mathbf{P}\cdot\mathbf{P})\boldsymbol{A}\Big] \\
+&\frac{1}{\epsilon}\Big[\lambda\boldsymbol{A} + 2\mathrm{i}\boldsymbol{A}_t\mathbf{P}\cdot\mathbf{Q}_t - \mathrm{i}\rho\boldsymbol{A}(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\cdot\mathbf{Q}_t \\
&\quad - \mathrm{i}(\lambda+\mu)\partial_{\mathbf{z}} : (Z^{-\mathrm{T}}\boldsymbol{A}\mathbf{P}) - \mathrm{i}(\lambda+\mu)\partial_{\mathbf{z}} : (\boldsymbol{A}\cdot\mathbf{P}Z^{-\mathrm{T}}) \\
&\quad + \partial_{\mathbf{z}} : (Z^{-\mathrm{T}}\partial_{\mathbf{Q}}\rho\boldsymbol{A}(\mathbf{P}\cdot\mathbf{Q}_t)^2) - 2\mu\mathrm{i}\partial_{\mathbf{z}} : (Z^{-\mathrm{T}}\mathbf{P}\boldsymbol{A}) \\
&\quad - \rho\boldsymbol{A}\,\mathrm{tr}\left((\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\otimes(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)Z^{-1}\partial_{\mathbf{z}}\mathbf{Q}\right) \\
&\quad - (\lambda+\mu)Z^{-1} : (\partial_{\mathbf{z}}\mathbf{Q}\boldsymbol{A}) - \mu\boldsymbol{A}\,\mathrm{tr}(Z^{-1}\partial_{\mathbf{z}}) \\
&\quad - \frac{1}{2}\partial^2_{\mathbf{QQ}}\rho\boldsymbol{A}(\mathbf{P}\cdot\mathbf{Q}_t)^2\,\mathrm{tr}(Z^{-1}\partial_{\mathbf{z}})\Big] \\
\epsilon^0&\Big[\rho\boldsymbol{A}_{tt} - \rho\boldsymbol{A}_t\mathbf{P}\cdot\mathbf{Q}_t - \epsilon\rho\partial_{\mathbf{z}} : (\rho Z^{-\mathrm{T}}M_1) \\
&\quad + \partial_{\mathbf{z}} : (Z^{-\mathrm{T}}\partial_{\mathbf{Q}}\rho(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\cdot\mathbf{Q}_t)\boldsymbol{A}) \\
&\quad + \partial_{\mathbf{z}} : (\partial_{\mathbf{z}} : (\rho\boldsymbol{A}(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\otimes(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)Z^{-1})Z^{-1}) \\
&\quad - (\lambda+\mu)\partial_{\mathbf{z}} : (\partial_{\mathbf{z}} : (\partial_{\mathbf{z}}(\mathbf{Q}A)Z^{-1})Z^{-1}) \\
&\quad - \mu\partial_{\mathbf{z}} : (\partial_{\mathbf{z}} : (Z^{-1})\boldsymbol{A}Z^{-1}) + \epsilon\partial_{\mathbf{Q}}\rho M_1\,\mathrm{tr}(Z^{-1}\partial_{\mathbf{z}}) \\
&\quad - \mathrm{i}\frac{1}{2}\partial^2_{\mathbf{QQ}}\rho\boldsymbol{A}(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\cdot\mathbf{Q}_t\,\mathrm{tr}(Z^{-1}\partial_{\mathbf{z}}) \\
&\quad - \partial_{\mathbf{z}} : (\partial_{\mathbf{z}} : (Z^{-1}\partial^2_{\mathbf{QQ}}\rho)\boldsymbol{A}(\mathbf{P}\cdot\mathbf{Q}_t)^2Z^{-1}) \\
&\quad + \partial_{\mathbf{z}} : \left(\mathrm{tr}(\partial^2_{\mathbf{QQ}}\rho Z^{-t}\partial_{\mathbf{z}}\mathbf{Q})(\mathbf{P}\cdot\mathbf{Q}_t)\left[(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\right]\boldsymbol{A}\right)\Big]
\end{aligned}
$$

Now define the operators $\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2$ acting on $\boldsymbol{A}$

$$
\mathcal{L}_0(\boldsymbol{A}) := -\rho\boldsymbol{A}(\mathbf{P}\cdot\mathbf{Q}_t)^2 + (\lambda+\mu)(\boldsymbol{A}\cdot\mathbf{P})\mathbf{P} + \mu(\mathbf{P}\cdot\mathbf{P})\boldsymbol{A} \tag{2.92}
$$

$$\begin{aligned}\mathcal{L}_1(\boldsymbol{A}) := \lambda\boldsymbol{A} + 2\mathrm{i}\boldsymbol{A}_t\mathbf{P}\cdot\mathbf{Q}_t - \mathrm{i}\rho\boldsymbol{A}(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\cdot\mathbf{Q}_t \\ - \mathrm{i}(\lambda+\mu)\partial_{\mathbf{z}} : (Z^{-\mathrm{T}}\boldsymbol{A}\mathbf{P}) - \mathrm{i}(\lambda+\mu)\partial_{\mathbf{z}} : (\boldsymbol{A}\cdot\mathbf{P}Z^{-\mathrm{T}}) \\ + \partial_{\mathbf{z}} : (Z^{-\mathrm{T}}\partial_{\mathbf{Q}}\rho\boldsymbol{A}(\mathbf{P}\cdot\mathbf{Q}_t)^2) - 2\mu\mathrm{i}\partial_{\mathbf{z}} : (Z^{-\mathrm{T}}\mathbf{P}\boldsymbol{A}) \\ - \rho\boldsymbol{A}\,\mathrm{tr}\left((\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\otimes(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)Z^{-1}\partial_{\mathbf{z}}\mathbf{Q}\right) \\ - (\lambda+\mu)Z^{-1} : (\partial_{\mathbf{z}}\mathbf{Q}\boldsymbol{A}) - \mu\boldsymbol{A}\,\mathrm{tr}(Z^{-1}\partial_{\mathbf{z}}) \\ - \frac{1}{2}\partial^2_{\mathbf{QQ}}\rho\boldsymbol{A}(\mathbf{P}\cdot\mathbf{Q}_t)^2\,\mathrm{tr}(Z^{-1}\partial_{\mathbf{z}}) \quad (2.93)\end{aligned}$$

$$\begin{aligned}\mathcal{L}_2(\boldsymbol{A}) := \rho\boldsymbol{A}_{tt} - \rho\boldsymbol{A}_t\mathbf{P}\cdot\mathbf{Q}_t - \epsilon\rho\partial_{\mathbf{z}} : (\rho Z^{-\mathrm{T}}M_1) \\ + \partial_{\mathbf{z}} : (Z^{-\mathrm{T}}\partial_{\mathbf{Q}}\rho(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\cdot\mathbf{Q}_t)\boldsymbol{A}) \\ + \partial_{\mathbf{z}} : (\partial_{\mathbf{z}} : (\rho\boldsymbol{A}(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\otimes(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)Z^{-1})Z^{-1})) \\ - (\lambda+\mu)\partial_{\mathbf{z}} : (\partial_{\mathbf{z}} : (\partial_{\mathbf{z}}(\mathbf{Q}A)Z^{-1})Z^{-1}) \\ - \mu\partial_{\mathbf{z}} : (\partial_{\mathbf{z}} : (Z^{-1})\boldsymbol{A}Z^{-1}) + \epsilon\partial_{\mathbf{Q}}\rho M_1\,\mathrm{tr}(Z^{-1}\partial_{\mathbf{z}}) \\ - \mathrm{i}\frac{1}{2}\partial^2_{\mathbf{QQ}}\rho\boldsymbol{A}(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\cdot\mathbf{Q}_t\,\mathrm{tr}(Z^{-1}\partial_{\mathbf{z}}) \\ - \partial_{\mathbf{z}} : (\partial_{\mathbf{z}} : (Z^{-1}\partial^2_{\mathbf{QQ}}\rho)\boldsymbol{A}(\mathbf{P}\cdot\mathbf{Q}_t)^2Z^{-1}) \\ + \partial_{\mathbf{z}} : \left(\mathrm{tr}(\partial^2_{\mathbf{QQ}}\rho Z^{-t}\partial_{\mathbf{z}}\mathbf{Q})(\mathbf{P}\cdot\mathbf{Q}_t)\left[(\mathbf{P}_t - \mathrm{i}\mathbf{Q}_t)\right]\boldsymbol{A}\right) \quad (2.94)\end{aligned}$$

Now $(\partial_t^2 - \mathcal{L})\mathbf{u}_{\mathrm{F}}$ can be written as

$$\begin{aligned}(\partial_t^2 - \mathcal{L})\mathbf{u}_{\mathrm{F}} = (2\pi\epsilon)^{-3/2}\sum_n \int_{\mathbb{R}^{3d}} (\epsilon^{-2}\mathcal{L}_{n,0}(\mathbf{a}_{n,0} + \epsilon\mathbf{a}_{n,1}) \\ \epsilon^{-1}\mathcal{L}_{n,1}(\mathbf{a}_{n,0} + \epsilon\mathbf{a}_{n,1}) + \mathcal{L}_{n,2}(\mathbf{a}_{n,0} + \epsilon\mathbf{a}_{n,1}))G_n^{\epsilon}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} \quad (2.95)\end{aligned}$$

Substituting the dynamics for $\mathcal{L}_{n,0}$ reveals that that $\mathcal{L}_{n,0}(\mathbf{a}_{n,0}) = 0$. Considering the $\mathcal{O}(1/\epsilon)$ term and equating to zero gives

$$\mathcal{L}_{n,1}(\mathbf{a}_{n,0}) = -\mathcal{L}_{n,0}(\mathbf{a}_{n,1}) \tag{2.96}$$

Now $\mathcal{L}_{n,0}$ is defined as

$$\mathcal{L}_{n,0} = \big(\mu|\mathbf{P}_n|^2 - \rho(\mathbf{P}_n \cdot \partial_t \mathbf{Q}_n)^2\big)\mathrm{Id}_3 + (\lambda+\mu)\mathbf{P}_n \otimes \mathbf{P}_n \tag{2.97}$$

which is a symmetric matrix with eigenvalues

$$\beta_{n,1} = (\lambda + 2\mu)|\mathbf{P}_n|^2 - \rho|\mathbf{P} \cdot \partial_t \mathbf{Q}_n|^2 \tag{2.98}$$

$$\beta_{n,2} = \mu|\mathbf{P}|^2 - \rho|\mathbf{P}_n \cdot \partial_t \mathbf{Q}_n|^2 \tag{2.99}$$

$$\beta_{n,3} = \mu|\mathbf{P}|^2 - \rho|\mathbf{P}_n \cdot \partial_t \mathbf{Q}_n|^2 \tag{2.100}$$

the corresponding eigenvectors are,

$$\mathbf{P}_n = (p_{n,1}, p_{n,2}, p_{n,3}) \tag{2.101}$$

$$\boldsymbol{d}_{n,1} = (-p_{n,2}, p_{n,1}, 0) \tag{2.102}$$

$$\boldsymbol{d}_{n,2} = (-p_{n,3}, 0, p_{n,1}). \tag{2.103}$$

For the P-wave, $n = \mathrm{p}$, taking inner product of with the eigenvectors;

$$\langle \mathbf{P}_\mathrm{p}, \mathcal{L}_{\mathrm{p},0}(\mathbf{a}_{n,1})\rangle = -\langle \mathbf{P}_\mathrm{p}, \mathcal{L}_{\mathrm{p},1}(\mathbf{a}_{\mathrm{p},0})\rangle \tag{2.104}$$

This gives,

$$\langle \mathcal{L}^*_{\mathrm{p},0}(\mathbf{P}_\mathrm{p}), \mathbf{a}_{n,1}\rangle = \langle \mathcal{L}_{\mathrm{p},0}(\mathbf{P}), \mathbf{a}_{\mathrm{p},1}\rangle$$
$$= ((\lambda+2\mu)|\mathbf{P}_\mathrm{p}|^2 - \rho|\mathbf{P}_\mathrm{p}\cdot\partial_t\mathbf{Q}_\mathrm{p}|^2)\langle \mathbf{P}_\mathrm{p}, \mathbf{a}_{\mathrm{p},1}\rangle = 0 \quad (2.105)$$

after plugging in the dynamics so we can recover the equation (2.34), as

$$\langle \mathbf{P}_\mathrm{p}, \mathcal{L}_{\mathrm{p},1}(\mathbf{a}_{\mathrm{p},0})\rangle = 0. \quad (2.106)$$

Considering $\boldsymbol{d}_{1,2}$,

$$\langle \boldsymbol{d}_{1,2}, \mathcal{L}_{\mathrm{p},0}(\mathbf{a}_{n,1})\rangle = -\langle \boldsymbol{d}_{1,2}, \mathcal{L}_{\mathrm{p},1}(\mathbf{a}_{\mathrm{p},0})\rangle. \quad (2.107)$$

Then

$$\langle \mathcal{L}^*_{\mathrm{p},0}(\boldsymbol{d}_{1,2}), \mathbf{a}_{\mathrm{p},1}\rangle = \langle \mathcal{L}_{\mathrm{p},0}(\boldsymbol{d}_{1,2}), \mathbf{a}_{\mathrm{p},1}\rangle = (\mu|\mathbf{P}|^2 - \rho|\mathbf{P}\cdot\partial_t\mathbf{Q}|^2)\langle \boldsymbol{d}_{1,2}, \mathbf{a}_{\mathrm{p},1}\rangle \quad (2.108)$$

And so plugging in the flow

$$\langle \boldsymbol{d}_{1,2}, \mathbf{a}_{\mathrm{p},1}\rangle = \frac{1}{\rho(c_\mathrm{s}^2 - c_\mathrm{p}^2)|\mathbf{P}|^2}\langle \boldsymbol{d}_{1,2}, \mathcal{L}_{\mathrm{p},1}(\mathbf{a}_{\mathrm{p},0})\rangle \quad (2.109)$$

Define the pseudo-inverse, for $\mathbf{v} \in \mathcal{S}(\mathbb{R}^3)$

$$\mathcal{L}^{-1}_{\mathrm{p},0}(\mathbf{v}) = \frac{1}{\rho(c_\mathrm{s}^2 - c_\mathrm{p}^2)|\mathbf{P}|^2}\left(\langle \hat{\boldsymbol{d}}_1, \mathbf{v}\rangle\hat{\boldsymbol{d}}_1 + \langle \hat{\boldsymbol{d}}_2, \mathbf{v}\rangle\hat{\boldsymbol{d}}_2\right) \quad (2.110)$$

and define

$$\mathbf{a}^\perp_\mathrm{p}\mathbf{v} = \mathcal{L}^{-1}_{\mathrm{p},0}\big((\mathrm{Id} - \Pi_\mathrm{p})\mathcal{L}_{\mathrm{p},1}(\mathbf{v})\big) \quad (2.111)$$

Where Π_{p} is projection onto $\mathbf{P}_{\mathrm{p}}$.

For the S-wave, $n = \mathrm{sv}, \mathrm{sh}$. From (2.96) we have

$$\mathcal{L}_{\mathrm{s},1}(\mathbf{a}_{\mathrm{s}}) = -\mathcal{L}_{\mathrm{s},0}(\mathbf{a}_{\mathrm{sh},0} + \mathbf{a}_{\mathrm{sv},0}) \tag{2.112}$$

Let $\boldsymbol{d}_{s,1} = \hat{\boldsymbol{N}}_{\mathrm{sh}}$, taking inner product with (2.112) gives

$$\langle \mathcal{L}_{\mathrm{s},0}(\hat{\boldsymbol{N}}_{\mathrm{sv}}), \mathbf{a}_{\mathrm{s},1} \rangle = (\mu |\mathbf{P}_{\mathrm{s}}|^2 - \rho |\mathbf{P}_{\mathrm{s}} \cdot \partial_t \mathbf{Q}_{\mathrm{s}}|^2) \langle \hat{\boldsymbol{N}}_{\mathrm{sv}}, \mathbf{a}_{\mathrm{s},1} \rangle = 0, \tag{2.113}$$

which is zero when the dynamics are substituted. From this we can get

$$\langle \hat{\boldsymbol{N}}_{\mathrm{sv}}, \mathcal{L}_{\mathrm{s},1} \mathbf{a}_{\mathrm{sv},0} \rangle = -\langle \hat{\boldsymbol{N}}_{\mathrm{sv}}, \mathcal{L}_{\mathrm{s},1} \mathbf{a}_{\mathrm{sh},0} \rangle, \tag{2.114}$$

Which gives us equation (2.35). The diabetic coupling term can be realized from the right hand side of (2.114) as $\langle \mathbf{a}_{\mathrm{sh},0}, \hat{\boldsymbol{N}}_{\mathrm{sv}} \rangle = 0$. Equation (2.36) can be recovered in a similar manor. Taking inner product with $\mathbf{P}_{\mathrm{s}}$ of (2.112) leads to,

$$\langle \mathbf{P}, \mathbf{a}_{\mathrm{s},1} \rangle = -\frac{1}{(\lambda + \mu)|\mathbf{P}|^2} \langle \mathbf{P}, \mathcal{L}_{\mathrm{s},1}(\mathbf{a}_{\mathrm{sv},0} + \mathbf{a}_{\mathrm{sh},0}) \rangle \tag{2.115}$$

Define the pseudo-inverse for $\mathbf{v} \in \mathcal{S}(\mathbb{R}^3)$

$$\mathcal{L}_{\mathrm{s},0}^{-1}(\mathbf{v}) = -\frac{1}{(\lambda + \mu)|\mathbf{P}_{\mathrm{s}}|^2} \langle \hat{\mathbf{P}}_{\mathrm{s}}, \mathbf{v} \rangle \hat{\mathbf{P}}_{\mathrm{s}} \tag{2.116}$$

and define

$$\mathbf{a}_{\mathrm{s}}^{\perp} \mathbf{v} = \mathcal{L}_{\mathrm{s},0}^{-1}\big((\mathrm{Id} - \Pi_{\mathrm{s}}) \mathcal{L}_{\mathrm{s},1}(\mathbf{v})\big). \tag{2.117}$$

with Π_{s} a projection onto the span of $\boldsymbol{d}_{\mathrm{s},1}$ and $\boldsymbol{d}_{\mathrm{s},2}$.

Definition 2.4.1 *Define the scaled semi-norm;*

$$\|\mathbf{u}(t,\cdot)\|_{\mathrm{E}} = \epsilon(\|\partial_t \mathbf{u}(t,\cdot)\|_{\mathrm{L}^2} + \|\nabla\cdot \mathbf{u}(t,\cdot)\|_{\mathrm{L}^2} + \|\nabla\times \mathbf{u}(t,\cdot)\|_{\mathrm{L}^2}) \tag{2.118}$$

Remark 1 *The scaling on the energy norm is needed as* $\partial_t \mathbf{u}, \nabla\cdot\mathbf{u}$ *and* $\nabla\times\mathbf{u}$ *are all* $\mathcal{O}(\epsilon^{-1})$*. This gives* $\|\mathbf{u}\|_{\mathbb{E}} = \mathcal{O}(1)$*.*

Proposition 2.4.2 *Let* $\mathbf{a}_{\mathrm{s}} = a_{\mathrm{sv}}\alpha_{\mathrm{sv}}\hat{\boldsymbol{N}}_{\mathrm{sv}} + a_{\mathrm{sh}}\alpha_{\mathrm{sh}}\hat{\boldsymbol{N}}_{\mathrm{sh}}$ *and* $\mathbf{a}_{\mathrm{p}} = a_{\mathrm{p}}\alpha_{\mathrm{p}}\hat{\boldsymbol{N}}_{\mathrm{p}}$*. The terms* $\mathbf{a}_{\mathrm{p}}$*,* $\mathbf{a}_{\mathrm{s}}$ *are bounded in the* L^2 *sense; furthermore,*

$$\|u_{\mathrm{F},1} - u_{\mathrm{F},0}\|_{\mathrm{E}} \leq \epsilon C_{T,\delta} \tag{2.119}$$

Proof: First we remark

$$\|\mathbf{a}_n(t,\cdot)\|_{\mathrm{L}^2} \leq \|\alpha_n\|_{\mathrm{L}^2}\|a_n(t,\cdot)\|_{\mathrm{L}^\infty} \text{ and } \|\mathbf{a}_n(t,\cdot)\|_{\mathrm{L}^\infty} \lesssim \|a_n(t,\cdot)\|_{\mathrm{L}^\infty} \tag{2.120}$$

From the definitions we have an immediate bound

$$\|u_{\mathrm{F},1}(t,\cdot) - u_{\mathrm{F},0}(t,\cdot)\|_{\mathrm{E}} \leq (2\pi\epsilon)^{-3d/2}\sum_n \epsilon\| \int_{\mathbb{R}^{3d}} \mathbf{a}_{n,1}^{\perp} + \mathbf{a}_{n,1}G_n \,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}\|_{\mathrm{E}} \tag{2.121}$$

Applying the derivatives with proposition 1.4.4, we have the bound

$$\|u_{\mathrm{F},1}(t,\cdot) - u_{\mathrm{F},0}(t,\cdot)\|_{\mathrm{E}} \leq \epsilon C\sum_n \|\mathbf{a}_{n,1}^{\perp}(t,\cdot) + \mathbf{a}_{n,1}(t,\cdot)\|_{\mathrm{L}^\infty} \tag{2.122}$$

The estimate of (2.121) then follows directly from Proposition 3.7 in [59]. We need to bound the prefactor terms, we note that on the compact set K_δ the bound for the prefactor terms fall from Lemma 5.4 in [59], We go through several of the bounds here, starting with the P-wave and dropping the subscripts as the calculations are the same

and setting $\mathbf{P} = \mathbf{P}_p$, $\mathbf{Q} = \mathbf{Q}_p$,

$$\partial_t a_{\mathrm{p},0} = a_{\mathrm{p},0}\left(\frac{\partial_{\mathbf{Q}} c_{\mathrm{p}} \cdot \mathbf{P}}{|\mathbf{P}|} + \frac{1}{2}\operatorname{tr}\left(Z^{-1}\partial_t Z\right)\right) \tag{2.123}$$

$$\partial_t a_{\mathrm{p},1} = a_{\mathrm{p},1}\left(\frac{\partial_{\mathbf{Q}} c_{\mathrm{p}} \cdot \mathbf{P}}{|\mathbf{P}|} + \frac{1}{2}\operatorname{tr}\left(Z^{-1}\partial_t Z\right)\right) + F_p(a_{\mathrm{p},0}, \mathbf{Q}, \mathbf{P}, c_{\mathrm{p}}) \tag{2.124}$$

With F being a smooth function in its arguments for $P, Q \in K_{\delta_T}$. Equation (2.123) immediately implies;

$$\partial_t |a_{\mathrm{p},0}| \leq |a_{\mathrm{p},0}| \left|\frac{\partial_{\mathbf{Q}} c_{\mathrm{p}} \cdot \mathbf{P}}{|\mathbf{P}|} + \frac{1}{2}\operatorname{tr}\left(Z^{-1}\partial_t Z\right)\right| \tag{2.125}$$

an application of Gronwalls gives,

$$\sup_{t\in[0,T]} \Lambda_{0,K_{\delta_T}}\left(a_{\mathrm{p},0}(t,\mathbf{q},\mathbf{p})\right) \leq C_{T,\delta} \tag{2.126}$$

To bound Equation (2.124) $\partial_{\mathbf{z}} a_{\mathrm{p},0}$ needs to be bounded, but Equation (2.125) shows.

$$\partial_t |\partial_{\mathbf{z}} a_{\mathrm{p},0}| \leq |\partial_{\mathbf{z}} a_{\mathrm{p},0}| \left|\left(\frac{\partial_{\mathbf{Q}} c_{\mathrm{p}} \cdot \mathbf{P}}{|\mathbf{P}|} + \frac{1}{2}\operatorname{tr}\left(Z^{-1}\partial_t Z\right)\right)\right| + |a_{\mathrm{p},0}| \left|\partial_{\mathbf{z}}\left(\frac{\partial_{\mathbf{Q}} c_{\mathrm{p}} \cdot \mathbf{P}}{|\mathbf{P}|} + \frac{1}{2}\operatorname{tr}\left(Z^{-1}\partial_t Z\right)\right)\right| \tag{2.127}$$

Then Gronwalls gives,

$$\sup_{t\in[0,T]} \Lambda_{1,K_{\delta_T}}\left(a_{\mathrm{p},0}(t,\mathbf{q},\mathbf{p})\right) \leq C_{\delta,T} \tag{2.128}$$

The function $F(a_{\mathrm{p},0}, \mathbf{Q}, \mathbf{P}, c_{\mathrm{p}})$ is differentiable with differentiable arguments on the compact set K_{δ_T}. Combining these we have

$$\sup_{t\in[0,T]} \Lambda_{\{0,1\},K_{\delta_T}}\left(a_{\mathrm{p},1}(t,\mathbf{q},\mathbf{p})\right) \leq C_{T,\delta} \tag{2.129}$$

Show it is shown that $\|a_{\mathrm{p}}\mathbf{P}\|_{\mathrm{L}^2}$ is bounded on $[0,T]\times K_{\delta/2}$. For the S-wave terms, again using the short notation $\mathbf{P}=\mathbf{P}_s$, $\mathbf{Q}=\mathbf{Q}_s$ and dropping the brach subscript, we can write the system ((2.35), (2.36)) as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\mathbf{a}_{\pm}^{sv}\\ \mathbf{a}_{\pm}^{sh}\end{pmatrix}=\frac{1}{2}\begin{pmatrix}h_{\pm} & m_{\pm}\\ -m_{\pm} & h_{\pm}\end{pmatrix}\begin{pmatrix}\mathbf{a}_{\pm}^{sv}\\ \mathbf{a}_{\pm}^{sh}\end{pmatrix} \tag{2.130}$$

where $m_{\pm}=\partial_t\hat{\boldsymbol{N}}_{\mathrm{sh}}\cdot\hat{\boldsymbol{N}}_{\mathrm{sv}}$ and

$$h_{\pm}=2\partial_{\mathbf{Q}_{\mathrm{s},\pm}}c_s\cdot\hat{\boldsymbol{N}}_{\mathrm{sh}}+\mathbf{a}^s\,\mathrm{tr}\left(\mathcal{Z}^{-1}\partial_t\mathcal{Z}^s\right) \tag{2.131}$$

Denote M as the matrix in eq. (2.130), and $\mathbf{a}=(\mathbf{a}^{sv},\mathbf{a}^{sh})^T$. Then the system can be recast as

$$\frac{\mathrm{d}\mathbf{a}}{\mathrm{d}t}=M(t)\mathbf{a} \tag{2.132}$$

Solving for the eigenvalues:

$$\lambda_{\mathrm{sh,sv}}(t)=-\partial_{\boldsymbol{Q}}c_{\mathrm{s}}\cdot\hat{\boldsymbol{N}}_{\mathrm{sh,sv}}-\frac{1}{2}\,\mathrm{tr}\Big(Z_{\mathrm{s}}^{-1}\frac{\mathrm{d}Z_{\mathrm{s}}}{\mathrm{d}t}\Big)\mp\mathrm{i}\frac{\mathrm{d}\hat{\boldsymbol{N}}_{\mathrm{sv,sv}}}{\mathrm{d}t}\cdot\hat{\boldsymbol{N}}_{\mathrm{sh,sv}} \tag{2.133}$$

To see that these are bounded simply note that form a smooth $\{\hat{N}_{\mathrm{p}},\hat{N}_{\mathrm{sh}},\hat{N}_{\mathrm{sh}}\}$ form an orthonormal frame and hence the last term in (2.133) is bounded for all $t\geq 0$.

$$\mathrm{tr}\Big(Z_{\mathrm{s}}^{-1}\frac{\mathrm{d}Z_{\mathrm{s}}}{\mathrm{d}t}\Big)=\frac{1}{\det(Z_{\mathrm{s}})}\frac{\mathrm{d}\det(Z_{\mathrm{s}})}{\mathrm{d}t}, \tag{2.134}$$

then by (2.1.5) we have a bound for $\det(Z_{\mathrm{s}})$ so Eq. (2.134) is bounded for all $t\geq 0$ and

$$\frac{\partial_{\mathbf{q}}H_{\mathrm{s}}\cdot\partial_{\mathbf{p}}H_{\mathrm{s}}}{H_{\mathrm{s}}}=-\frac{\partial_{\boldsymbol{Q}_{\mathrm{s}}}c_{\mathrm{s}}\cdot\boldsymbol{P}_{\mathrm{s}}}{|\boldsymbol{P}_{\mathrm{s}}|} \tag{2.135}$$

by observation, with (2.16), Eq. (2.135) is bounded for all $t \geq 0$. The the eigenvalues in Eq. (2.133) are bounded for all $t \geq 0$. So we have the

$$\sup_{t\in[0,T]} \Lambda_{0,K_{\delta/2}}\big(a_{\mathrm{s},0}(t,\mathbf{q},\mathbf{p})\big) \leq C_{\delta,T} \tag{2.136}$$

For $\partial_t \mathbf{a}_{\mathrm{s},1}$, we have the system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} a_{\mathrm{sv},1} \\ a_{\mathrm{sh},1} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} h & m \\ -m & h \end{pmatrix}\begin{pmatrix} a_{\pm}^{\mathrm{sv}} \\ a_{\pm}^{\mathrm{sh}} \end{pmatrix} + \mathbf{F}(\mathbf{a}_{\mathrm{s},0}, \mathbf{Q}, \mathbf{P}, c_{\mathrm{s}}) \tag{2.137}$$

Then the bounds follow from previous work, as $\mathbf{F}$ is smooth away from $|\mathbf{P}| > 0$, so we arrive at

$$\sup_{t\in[0,T]} \Lambda_{1,K_{\delta/2}}\big(a_{\mathrm{s},1}(t,\mathbf{q},\mathbf{p})\big) \leq C_{\delta,T} \tag{2.138}$$

■

Definition 2.4.3 *Define a standard smooth cutoff function* $\chi_\delta : \mathbb{R}^{2d} \to [0,1]$ *for the set* K_δ *as*

$$\chi_\delta(\mathbf{q},\mathbf{p}) = \begin{cases} 1, & (\mathbf{q},\mathbf{p}) \in K_\delta \\ 0, & (\mathbf{q},\mathbf{p}) \in \mathbb{R}^d \backslash K_\delta \end{cases} \tag{2.139}$$

and for any $k \in \mathbb{N}$, *there exists a constant* $C_{K,\delta}$ *such that*

$$\Lambda_k\big(\chi_\delta(\mathbf{q},\mathbf{p})\big) < C_{K,\delta} \tag{2.140}$$

Definition 2.4.4 *We define the filtered version of the FGA as follows;*

$$\tilde{u}_{\mathrm{F},1}(t,\mathbf{x}) = (2\pi\epsilon)^{-3d/2}\int_{\mathbb{R}^{3d}} \chi_\delta \sum_{j,n=0}^{1,5} \epsilon^j \big(\mathbf{a}_{n,j}(t,\mathbf{y},\mathbf{q},\mathbf{p}) + \mathbf{a}^{\perp}_{n,j}(\mathbf{a}_{n,0}(t,\mathbf{y},\mathbf{q},\mathbf{p}))\big) \times G^{\epsilon}_n(t,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{q} \quad (2.141)$$

where $\mathbf{a}^{\perp}_{n,0} = 0$ *and* $\mathbf{a}^{\perp}_{n,1}$ *are defined from* (2.111), (2.117).

Proposition 2.4.5 *For any* $T > 0$ *and* $t \in [0,T]$

$$\|\mathbf{u}_{\mathrm{F},1}(t,\cdot) - \tilde{\mathbf{u}}_{\mathrm{F},1}(t,\cdot)\|_{\mathrm{E}} = \mathcal{O}(\epsilon^\infty) \quad (2.142)$$

Proof: Starting from the definition,

$$\begin{aligned}
&\|\mathbf{u}_{\mathrm{F},1} - \tilde{\mathbf{u}}_{\mathrm{F},1}\|_{\mathrm{E}} && (2.143)\\
&\leq (2\pi\epsilon)^{-3d/2}\int_{\mathbb{R}^{3d}} \Big\| \sum_n (1-\chi_\delta)(\mathbf{a}_{n,0} + \epsilon\mathbf{a}_{n,1}) e^{\frac{\mathrm{i}}{\epsilon}\Phi_n}\Big\|_{\mathrm{E}}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{q}\\
&\leq 2^{-d/2}\sum_n \|(1-\chi_\delta)(a_{n,0}+\epsilon a_{n,1})\hat{\boldsymbol{N}}_n \mathcal{F}\alpha_n\|_{\mathrm{E}}\\
&\leq 2^{-d/2}\Big(\|\mathcal{F}^\epsilon u^\epsilon_0\|_{\mathrm{L}^2(\mathbb{R}^n\setminus K_\delta)} + \epsilon\|\mathcal{F}^\epsilon u^\epsilon_1\|_{\mathrm{L}^2(\mathbb{R}^n\setminus K_\delta)}\Big)\sum_n \|(1-\chi_\delta)(a_{n,0}+\epsilon a_{n,1})\hat{\boldsymbol{N}}_n\|_{\mathrm{L}^\infty}\\
&\leq C_{\delta,T}\Big(\|\mathcal{F}^\epsilon u^\epsilon_0\|_{\mathrm{E}(\mathbb{R}^n\setminus K_\delta)} + \epsilon\|\mathcal{F}^\epsilon u^\epsilon_1\|_{\mathrm{E}(\mathbb{R}^n\setminus K_\delta)}\Big) && (2.144)\\
&\leq C_{\delta,T}\Big(\|\mathcal{F}^\epsilon u^\epsilon_0\|_{\mathrm{L}^2(\mathbb{R}^n\setminus K_\delta)} + \epsilon\|\mathcal{F}^\epsilon u^\epsilon_1\|_{\mathrm{L}^2(\mathbb{R}^n\setminus K_\delta)}\Big) = \mathcal{O}(\epsilon^\infty) && (2.145)
\end{aligned}$$

Where the second equality is from Proposition (1.4.4), the third inequality is by similar arguments found in Proposition (2.4.2). Also from Eq. (2.40), with the direct bound

$$\|\alpha_n\|_{\mathrm{E}} \leq C(\|\mathbf{u}^\epsilon_0\|_{\mathrm{E}} + \epsilon\|\mathbf{u}^\epsilon_1\|_{\mathrm{E}}). \quad (2.146)$$

The last inequality is justified noticing the derivatives do not affect the initial conditions only G_n^ϵ in $\mathcal{F}^\epsilon u_0^\epsilon$ and $\mathcal{F}^\epsilon u_1^\epsilon$. ■

Proposition 2.4.6 *The operators $\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2$ are bounded. That is, for a given T and any $t \in [0,T]$, $\mathbf{a} \in C^\infty([0,T]) \times S(\mathbb{R}^{2d})$ and for $k = 0, 1$, $j = 0, 1, 2$*

$$\sup_{t\in[0,T]} \Lambda_{k,K_\delta}\big(\mathcal{L}_j(\mathbf{a})\big) < C_{T,K_\delta} \text{ and } \|\mathcal{L}_j(\mathbf{a}(t,\cdot))\|_{\mathrm{L}^\infty} < C_{T,\delta} \tag{2.147}$$

Proof: Notice $\mathcal{L}_{n,j}$ depend on $\mathbf{P}_n$, $\mathbf{Q}_n$ Z_n^{-1} it its derivatives, all of which are bounded on $[0,T] \times K_\delta$, this gives the result ■

Proposition 2.4.7 *For a given T and any $t \in [0,T]$, for $k = 0, 1$ and $\mathbf{a} \in C^\infty([0,T]) \times S(\mathbb{R}^{2d})$, we have*

$$\sup_{t\in[0,T]} \Lambda_{k,K_\delta}\big(\mathbf{a}_{n,1}^\perp(\mathbf{a}(t,\cdot))\big) < C_{T,K_\delta} \text{ and } \sup_{t\in[0,T]} \|\mathbf{a}_{n,1}^\perp(\mathbf{a}(t,\cdot))\|_{\mathrm{L}^\infty} < C_{T,\delta} \tag{2.148}$$

Proof: For both $\mathbf{a}_{\mathrm{p}}^\perp$ and $\mathbf{a}_{\mathrm{s}}^\perp$, the pseudo-operators $\mathcal{L}_{\mathrm{p},0}^{-1}$, and $\mathcal{L}_{\mathrm{s},0}^{-1}$ are bounded on K_δ as $|P| > 0$ as $|\mathcal{L}_{n,0}^{-1}|_{\mathrm{L}^\infty} \leq C\delta^{-1}$. Then by proposition 2.4.6 we the result. ■

Proposition 2.4.8 *Consider the Elastic wave equation with a forcing term.*

$$\begin{cases} \rho(x)\partial_t^2 \mathbf{u} - (\lambda + 2\mu)\nabla(\nabla\cdot \mathbf{u}) + \mu\, \nabla\times\nabla\times \mathbf{u} = \mathbf{F}(t,\mathbf{x}) & \\ u(0,\mathbf{x}) = \mathbf{u}_0^\epsilon, & \mathbf{x} \in \mathbb{R}^d\,. \\ u(t,\mathbf{x}) = \mathbf{u}_1^\epsilon, & \mathbf{x} \in \mathbb{R}^d \end{cases} \tag{2.149}$$

Let $T > 0$, and let $\mathbf{u}_0(t,\mathbf{x}) \in C^\infty([0,T]) \times H_0^1(\mathbb{R}^d)$. For each $t \in [0,T]$ Then we have the

following estimate:

$$\|\rho\partial_t \mathbf{u}\|_{\mathrm{L}^2} + \|(\lambda+2\mu)\,\nabla\cdot\mathbf{u}\|_{\mathrm{L}^2} + \|\mu\,\nabla\times\mathbf{u}\|_{\mathrm{L}^2} \leq C_T\left(\frac{1}{\epsilon}\|\mathbf{u}(0,\cdot)\|_{\mathrm{E}} + \int_0^t \|\mathbf{F}(s,\cdot)\|_{\mathrm{L}^2}\,\mathrm{d}\right) \tag{2.150}$$

In particular,

$$\sup_{t\in[0,T]} \|\mathbf{u}(t,\cdot)\|_{\mathrm{E}} \leq C_T\left(\|\mathbf{u}_0^\epsilon\|_{\mathrm{E}} + \epsilon\int_0^T \|\mathbf{F}(s,\cdot)\|_{\mathrm{L}^2}\,\mathrm{d}\right) \tag{2.151}$$

Proof: This is a standard estimate, dotting Eq. (2.150) with $\partial_t\mathbf{u}$ and integrating over space we have

$$\frac{\partial_t}{2}\int_{\mathbb{R}^d} \rho(x)|\partial_t\mathbf{u}|^2 + (\lambda+2\mu)|\,\nabla\cdot\mathbf{u}|^2 + \mu|\,\nabla\times\mathbf{u}|^2\,\mathrm{d}\mathbf{x} \leq \int_{\mathbb{R}^d} |\mathbf{u}_t\cdot\mathbf{F}|\,\mathrm{d}\mathbf{x} \tag{2.152}$$

The right-hand side can then be estimated by

$$\int_{\mathbb{R}^d} |\mathbf{u}_t\cdot\mathbf{F}|\,\mathrm{d}\mathbf{x} \leq \frac{1}{2}\int_{\mathbb{R}^d} |\partial_t\mathbf{u}|^2 + |\mathbf{F}|^2\,\mathrm{d}\mathbf{x} \tag{2.153}$$

adding the missing terms to apply Gronwall's give the bound

$$e^t(\|\rho(x)\mathbf{u}_1\|_{\mathrm{L}^2}^2 + \|(\lambda+\mu)\nabla\cdot\mathbf{u}_0\|_{\mathrm{L}^2}^2 + \|\mu\,\nabla\times\mathbf{u}\|_{\mathrm{L}^2}^2) + \int_0^t e^{t-s}\int_{\mathbb{R}^d}|\mathbf{F}(s,\mathbf{x})|^2\,\mathrm{d}\mathbf{x}\,\mathrm{d}s \tag{2.154}$$

Taking max over ρ,λ,μ and over T we arrive at the estimate ■

Proposition 2.4.9 *Let* $\mathbf{u}$ *solve the Cauchy problem* (2.2)*. If* $\mathbf{u}_{\mathrm{F},0}$ *is the first order FGA approximation* (2.52)*, then we have the following estimate with the initial conditions.*

$$\|\mathbf{u}(0,\mathbf{x}) - \mathbf{u}_{\mathrm{F}}(0,\mathbf{x})\|_{\mathrm{E}} \leq \epsilon C_T \tag{2.155}$$

Proof: First computing the following;

$$\partial_t \mathbf{a}_n(0,\mathbf{y},\mathbf{q},\mathbf{p}) = \alpha(\mathbf{y},\mathbf{q},\mathbf{p})\left(\frac{\partial c(\mathbf{q})\cdot\mathbf{p}}{|\mathbf{p}|} - d\right)\hat{\boldsymbol{n}}_n \tag{2.156}$$

For estimating $\mathbf{u}(0,\mathbf{x}) - \mathbf{u}_{\mathrm{F},0}(0,\mathbf{x})$ in the energy norm, we can write $\partial_t u(0,x) = u_1^\epsilon(\mathbf{x})$. In terms of the FIO this gives;

$$\begin{aligned}\int_{\mathbb{R}^{3d}} u_0^\epsilon(\mathbf{y})\frac{\mathrm{i}}{\epsilon}\Phi_n(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})G_n^\epsilon(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}& \\ = \int_{\mathbb{R}^{3d}} \mathbf{u}_1^\epsilon(\mathbf{y})G_n^\epsilon(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}.&\end{aligned} \tag{2.157}$$

Then

$$\begin{aligned}|\mathbf{u}_1^\epsilon(0,\mathbf{x}) - \partial_t\mathbf{u}_{\mathrm{F},0}(0,\mathbf{x})| = (2\pi\epsilon)^{-d/2}\Big|\sum_n \int_{\mathbb{R}^{3d}} \big(&\frac{1}{2}(\mathbf{u}_1^\epsilon(y)\cdot\hat{\boldsymbol{n}})\hat{\boldsymbol{n}} \\ &- \partial_t\mathbf{a}_n(0,\mathbf{y},\mathbf{q},\mathbf{p}) - \mathbf{a}_n\frac{\mathrm{i}}{\epsilon}\Phi_n(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\big)G_n^\epsilon(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}\end{aligned} \tag{2.158}$$

For one terms this is after plugging in (2.157),

$$\begin{aligned}\int_{\mathbb{R}^{3d}} \Big(u_0^\epsilon(\mathbf{y})\frac{\mathrm{i}}{2\epsilon}\Phi_n(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p}) - \alpha(\mathbf{y},\mathbf{q},\mathbf{p})\left(\frac{\partial c(\mathbf{q})\cdot\mathbf{p}}{|\mathbf{p}|} - d\right)\hat{\boldsymbol{n}}_n& \\ - 2^{d/2}\alpha(\mathbf{y},\mathbf{q},\mathbf{p})\hat{\boldsymbol{n}}_n\frac{\mathrm{i}}{\epsilon}\Phi_n(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\Big)& \\ \times G_n^\epsilon(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}&\end{aligned} \tag{2.159}$$

Plugging in α_n from (2.40) and summing over the wavefields and branches gives

$$\mathbf{u}_1^\epsilon(\mathbf{x}) - \partial_t \mathbf{u}_{\mathrm{F},0}(0,\mathbf{x}) = \\ -\sum_n \int_{\mathbb{R}^{3d}} \frac{1}{2c_n|\mathbf{p}|^3}\Big(\boldsymbol{u}_0^\epsilon(\boldsymbol{y})c_n|\boldsymbol{p}| \pm \mathrm{i}\epsilon \boldsymbol{u}_1^\epsilon(\boldsymbol{y})\Big)\cdot \hat{\boldsymbol{n}}_n \\ \times \left(\frac{\partial c(\mathbf{q})\cdot \mathbf{p}}{|\mathbf{p}|} - d\right)\hat{\boldsymbol{n}}_n G_n^\epsilon(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} \quad (2.160)$$

By Proposition (1.4.4) we can arrive at the estimate

$$\|\mathbf{u}_1^\epsilon - \partial_t \mathbf{u}_{\mathrm{F},0}(0,\cdot)\|_{\mathrm{L}^2} \leq C_T \quad (2.161)$$

For the Div term,

$$\nabla\cdot \mathbf{u}_{\mathrm{F}}(0,\mathbf{x})\,\mathrm{d}\mathbf{x} \\ = (2\pi\epsilon)^{-d/2}\sum_n \int_{\mathbb{R}^{3d}} \nabla\cdot\Big(\mathbf{a}_n(0,\mathbf{y},\mathbf{q},\mathbf{p})G_n^\epsilon(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\Big)\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} \\ = (2\pi\epsilon)^{-d/2}\sum_n \int_{\mathbb{R}^{3d}} \frac{\mathrm{i}}{\epsilon}\mathbf{a}_n\cdot(\mathbf{P}_n + (\mathbf{x}-\mathbf{Q}_n))\,G_n^\epsilon(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} \quad (2.162)$$

applying the operators and integration by parts gives

$$\int_{\mathbb{R}^d}\left(\frac{\mathrm{i}}{\epsilon}\mathbf{a}_n\cdot\mathbf{p}_n - \partial_{\mathbf{z}}(Z^{-1}\mathbf{a})\right)G_n^\epsilon(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\mathbf{x} \quad (2.163)$$

Now consider the difference, $\nabla\cdot\mathbf{u}_0^\epsilon - \nabla\cdot\mathbf{u}_{\mathrm{F},0}(t,\mathbf{x})$. Writing in terms of the FIO and writing one term

$$\int_{\mathbb{R}^{3d}}\left(\frac{\mathrm{i}}{\epsilon}\big(\mathbf{u}_0^\epsilon(\mathbf{y}) - \mathbf{a}(0,\mathbf{y},\mathbf{q},\mathbf{p})\big)\cdot\mathbf{P}_n + \partial_{\mathbf{z}}(Z^{-1}(\mathbf{u}_0^\epsilon(\mathbf{y}) - \mathbf{a}(0,\mathbf{y},\mathbf{q},\mathbf{p})))\right) \\ \times G_n^\epsilon(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} \quad (2.164)$$

With $\mathbf{a}(0,\mathbf{y},\mathbf{q},\mathbf{p}) = \alpha_n(\mathbf{y},\mathbf{q},\mathbf{p})\hat{\boldsymbol{n}}$ and summing over n we have;

$$\nabla\cdot\mathbf{u}_0^\epsilon(\mathbf{x}) - \nabla\cdot\mathbf{u}_{\mathrm{F},0}(0,\mathbf{x}) = \sum_n \int_{\mathbb{R}^d} \partial_{\mathbf{z}}(Z^{-1}(\mathbf{u}_0^\epsilon(\mathbf{y}) - \alpha_n(\mathbf{y},\mathbf{q},\mathbf{p})\hat{\boldsymbol{n}})G_n^\epsilon(0,\mathbf{x},\mathbf{y},\mathbf{q},\mathbf{p})\mathbf{x} \quad (2.165)$$

Again, by Proposition (1.4.4) we arrive at the estimate

$$\|\nabla\cdot\mathbf{u}_0^\epsilon - \nabla\cdot\mathbf{u}_{\mathrm{F},0}(t,\cdot)\,\mathrm{d}\mathbf{x}\|_{\mathrm{L}^2} \leq C_T \quad (2.166)$$

The Curl term has a similar estimate as the Div term. These 3 estimates show the result.

■

Theorem 2.4.10 *Let $\{u_0^\epsilon\}$ be a family of asymptotically high frequency initial conditions, and let $\mathbf{u}$ solve the Cauchy problem (2.2). If $\mathbf{u}_{\mathrm{F},0}$ is the first order FGA approximation (2.52), then for a given T and any $t \in [0,T]$, $\delta > 0$ and sufficiently small ϵ, we have*

$$\sup_{t\in[0,T]} \|\mathbf{u}(t,\cdot) - \mathbf{u}_{\mathrm{F},0}(t,\cdot)\|_{\mathrm{E}} \leq \epsilon C_{T,\delta} \quad (2.167)$$

Proof:

Proposition 2.4.11 *We have*

$$\|(\partial_t^2 - \mathcal{L})\tilde{\mathbf{u}}_{\mathrm{F},1}\|_{\mathrm{E}} \leq \epsilon C_{T,\delta} \quad (2.168)$$

Proof: Plugging $\tilde{\mathbf{u}}_{\mathrm{F},1}$ into (2.2) gives,

$$(\partial_t^2 - \mathcal{L})\tilde{\mathbf{u}}_{\mathrm{F},1}(t,\mathbf{x}) = (2\pi\epsilon)^{-3d/2}\int_{\mathbb{R}^{3d}} \chi_\delta \sum_{n,m} \epsilon^{m-2}\mathcal{L}_{n,m}\Big(\mathbf{a}_{n,0} + \epsilon(\mathbf{a}_{n,1} + \mathbf{a}_{n,1}^\perp(\mathbf{a}_{n,0}))\Big)G_n^\epsilon\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{q} \quad (2.169)$$

Expanding and simplifying yields;

$$
\begin{aligned}
&(\partial_t^2 - \mathcal{L})\tilde{\mathbf{u}}_{\mathrm{F},1}(t,\mathbf{x}) \\
&\qquad = (2\pi\epsilon)^{-3/2} \sum_n \chi_\delta G_n^\epsilon \,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} \int_{\mathbb{R}^{3d}} \epsilon^{-2}\mathcal{L}_{n,0}(\mathbf{a}_{n,0}) \\
&\quad + \epsilon^{-1}\mathcal{L}_{n,0}(\mathbf{a}_{n,1}) + \epsilon^{-1}\mathcal{L}_{n,1}(\mathbf{a}_{n,0}) + \mathcal{L}_{n,1}(\mathbf{a}_{n,1}) + \mathcal{L}_{n,2}(\mathbf{a}_{n,0} + \epsilon\mathbf{a}_{n,1}) \\
&\qquad - (\epsilon^{-1}\mathcal{L}_{n,0} + \mathcal{L}_{n,1} + \epsilon\mathcal{L}_{n,2})\mathcal{L}_{n,0}^{-1}((\mathrm{Id} - \Pi_n)(\mathcal{L}_{n,1}(\mathbf{a}_{n,0}\hat{\boldsymbol{N}}_n)))
\end{aligned}
\tag{2.170}
$$

Direct cancellation yields

$$
\begin{aligned}
&(\partial_t^2 - \mathcal{L})\tilde{\mathbf{u}}_{\mathrm{F},1}(t,\mathbf{x}) \\
&\qquad = (2\pi\epsilon)^{-3/2} \sum_n G_n^\epsilon \,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} \int_{\mathbb{R}^{3d}} R_0(t,\mathbf{p},\mathbf{q}) + \epsilon R_1(t,\mathbf{p},\mathbf{q})
\end{aligned}
\tag{2.171}
$$

where

$$
R_0(t,\mathbf{p},\mathbf{q}) = \mathcal{L}_{n,1}(\mathbf{a}_{n,1}) + \mathcal{L}_{n,2}(\mathbf{a}_{n,0}) - \mathcal{L}_{n,1}\mathbf{a}_{n,1}^{\perp}(\mathbf{a}_{n,0}) \tag{2.172}
$$

$$
R_1(t,\mathbf{p},\mathbf{q}) = \mathcal{L}_{n,2}(\mathbf{a}_{n,1}) - \mathcal{L}_{n,2}\mathbf{a}_{n,1}^{\perp}(\mathbf{a}_{n,0}) \tag{2.173}
$$

Then by Propositions 1.4.4 and 2.4.7

$$
\|(\partial_t^2 - \mathcal{L})\tilde{\mathbf{u}}_{\mathrm{F},1}(t,\cdot)\|_{\mathrm{L}^2} \leq C_{T,\delta}(\|R_0(t,\cdot)\|_{\mathrm{L}^\infty} + \epsilon\|R_1(t,\cdot)\|_{\mathrm{L}^\infty}) \tag{2.174}
$$

By prop. 2.4.7, 2.4.6 R_0 and R_1 are bounded. With the definition of the energy norm, (2.4.8) and (2.4.9) we have the result. ■

$$
\|\mathbf{u} - \mathbf{u}_{\mathrm{F},0}\|_{\mathrm{E}} \leq +\|\mathbf{u} - \tilde{\mathbf{u}}_{\mathrm{F},1}\|_{\mathrm{E}} + \|\tilde{\mathbf{u}}_{\mathrm{F},1} - \mathbf{u}_{\mathrm{F},1}\|_{\mathrm{E}} + \|\mathbf{u}_{\mathrm{F},1} - \mathbf{u}_{\mathrm{F},0}\|_{\mathrm{E}} \tag{2.175}
$$

For the first term define the quantity $e = \mathbf{u} - \tilde{\mathbf{u}}_{\mathrm{F},1}$. By prop. 2.4.8

$$\|e\|_{\mathrm{E}} \leq C_{T,\delta}\big(\|e(0,\cdot)\|_{\mathrm{E}} + \epsilon \int_0^t \|R_0\|_{\mathrm{L}^2}\,\mathrm{d}s\big) + \mathcal{O}(\epsilon^2) \tag{2.176}$$

Prop 2.4.9 shows that $\|e(0,\cdot)\|_{\mathrm{E}} \leq \epsilon C_{T,\delta}$. Also, prop. 2.4.7 and 2.4.6 show $\|R_0\|_{\mathrm{L}^2} \leq C_{T,\delta}$. Thus the first is estimated at the correct order. For the second term, is $\mathcal{O}(\epsilon^\infty)$ by 2.4.5, and the last term is estimated to the desired order by prop. 2.4.2. ■

As a consequence of the energy norm, we have the following corollary;

Corollary 1 *Under the same assumptions as Theorem 2.4.10, define U_{F} as*

$$U_{\mathrm{F}} = (\partial_t \mathbf{u}_{\mathrm{F}}, \nabla\cdot \mathbf{u}_{\mathrm{F}}, \nabla\times \mathbf{u}_{\mathrm{F}}) \tag{2.177}$$

Define X_{F} as in eq. (2.26) and let X be the solution to the hyperbolic system (2.5), then we have the estimate

$$\sup_{t\in[0,T]} \|X_{\mathrm{F},0}(t,\cdot) - U_{\mathrm{F},0}(t,\cdot)\|_{\mathrm{L}^2} \leq \epsilon C_{T,\delta} \tag{2.178}$$

Proof: Applying the triangle inequality gives the immediate result.

$$\|X_{\mathrm{F},0} - U_{\mathrm{F},0}\|_{\mathrm{L}^2} \leq \|X - X_{\mathrm{F},0}\|_{\mathrm{L}^2} + \|X - U_{\mathrm{F},0}\|_{\mathrm{L}^2} \leq \epsilon C_{T,\delta} + \|\mathbf{u} - \mathbf{u}_{\mathrm{F},0}\|_{\mathrm{E}} \tag{2.179}$$

Where $\|X - X_{\mathrm{F},0}\|_{\mathrm{L}^2}$ is bounded from (2.25). ■

Chapter 3

Frozen Gaussian Approximation and Seismology

This chapter contains a brief discussion on seismic tomography along with the derivation necessary tools for numerical experiments. These tools are interface and boundary conditions for reflection and transmission coefficients and the Greens function for a point source.

3.1 Remarks on Seismology and the FGA

Various geophysical aspects, e.g., tectonics and geodynamics [2, 77, 76, 106], can be better understood by images of substructures (e.g. locations of seismic interfaces) of the Earth generated by seismic tomography. Traveltime tomography [1, 108, 61, 70, 74, 3, 90] has been used extensively to approximate subsurface velocity profiles. The velocity model is updated iteratively by backwards propagating the traveltime residuals along rays. This method is fairly inexpensive and has been is a standard for determining the Earth's velocity profile. Generally for the forward modeling, a ray theory based solver is used.

In a frequency limit, ray theory is exact; however, all data collected is of finite frequency for which rays theory is approximate. In addition, traveltime tomography only takes the traveltime information into account while discarding amplitude information.

Full waveform inversion (FWI), is an inversion technique that uses the waveform to invert a velocity profiles. All the information contained from recorded wavefields is used; i.e, traveltime and amplitude information, thus FWI has higher accuracy and resolution then traveltime tomography. The FWI algorithm, like traveltime tomography, is formulated as an iterative descent method that for which the minimization of a residual is achieved. The gradient calculated at each iteration provides the search direction for the minimization of the object functional.

Compared to other ray-based methods including: WKBJ [13, 14], WKM [66, 40], generalized ray theory [39, 93], seismic traveltime tomography [1, 90], Kirchhoff migration [32, 50] and Gaussian beam migration [42, 43, 67, 33, 31, 69], the FGA does not need to solve ray paths by shooting to reach the receivers, and provides accurate solutions at the presence of caustics and multipathing, with no requirement on tuning beam width parameters to achieve a good resolution [10, 42, 26, 72, 58]. With a multicore processors computer station, the property that the FGA algorithm is embarrassingly parallel makes possible the application of FGA to compute 3-D high frequency sensitivity kernels, and further used for 3-D traveltime tomography and FWI. Despite the similarities, the GBM and FGA are different in nature. The FGA is based on phase plane analysis, while the GBM is based on the asymptotic solution with WKBJ initial data. The FGA overcomes beam spreading by fixing the beam width and the FGA exhibits better accuracy than the GBM. The K^{th} order GBM converges to the true solution with an accuracy of $\mathcal{O}(\epsilon^{K/2})$, while the K^{th} order FGA has a convergence order of $\mathcal{O}(\epsilon^{K})$ [59]. The FGA approximates

the solution to the wave equation by an integral representation, with dynamics In recent works [11, 12], the authors develop and verify the FGA as an efficient solver for computing high-frequency acoustic wave propagation. The central idea of FGA is to approximate wavefields by fixed-width Gaussian wave packets, whose dynamics follow ray paths with the prefactor amplitude equation derived from an asymptotic expansion on phase plane. The wavefield can be reconstructed using the inverse FBI due to a canonical transform, a transform that preserves the dynamics of a Hamiltonian system.

3.2 Interface and Boundary Conditions

We to apply the formulation in the local tangent-normal coordinates by treating the tangential direction as the local flat horizontal interface. For simplicity, let Γ denote a smooth interface that describes the graph of a function $z = f(x, y)$ with normal N. The wave speeds of the two layers are assumed to be,

$$c_{\mathrm{p}}(\boldsymbol{x}) = \begin{cases} c_{\mathrm{p}}^{\vee}(\boldsymbol{x}) & z > f(x,y) \\ c_{\mathrm{p}}^{\wedge}(\boldsymbol{x}) & z < f(x,y) \end{cases}, \quad c_{\mathrm{s}}(\boldsymbol{x}) = \begin{cases} c_{\mathrm{s}}^{\vee}(\boldsymbol{x}) & z > f(x,y) \\ c_{\mathrm{s}}^{\wedge}(\boldsymbol{x}) & z < f(x,y) \end{cases}. \tag{3.1}$$

In Fig. 3.1, we only consider an incident Gaussian wave packet for P-wave hitting the a flat interface located at $z = z_0$, and then reflected and transmitted as Gaussian wave packets for P- and SV-waves, respectively. The other cases including an incident Gaussian wave packet for SV- and SH-waves can be handled similarly, although there is no interaction between P- and S-waves for the case of SH-wave. Associated with each Gaussian wave packet, one needs to provide the reflection and transmission conditions for $a_{\mathrm{p,s}}$, $\boldsymbol{Q}_{\mathrm{p,s}}$, and $\boldsymbol{P}_{\mathrm{p,s}}$, which follow the Snell's Law and the Zoeppritz equations [103]. However, for FGA, what is different from standard ray theory on the flat interface is that, one also needs to

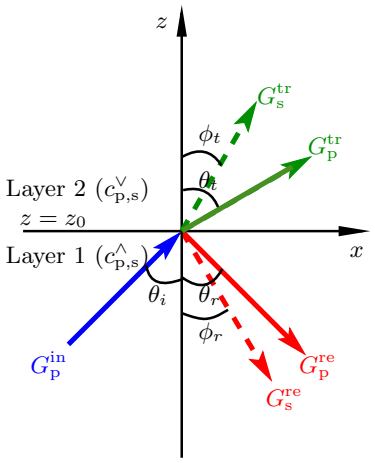


Figure 3.1: Cartoon illustration of an incident Gaussian wave packet for P-wave hitting the interface at $z = z_0$, and then reflected and transmitted as Gaussian wave packets for P- and SV-waves. Here the $G^{\mathrm{in,re,tr}}_{\mathrm{p,s}}$ stands for the Gaussian wave packet for the incident, reflected and transmitted P- and SV-waves, respectively. We denote $\theta_i, \theta_r, \theta_t$ to be the incident, reflection and transmission angles of P-waves, and ϕ_r, ϕ_t to be the reflection and transmission angles of SV-waves, respectively.

derive the interface conditions for $Z_{\mathrm{p,s}}$ which will change after the Gaussian wave packet hits the interface and affect the dynamics of $a_{\mathrm{p,s}}$ given by eqs. (2.34)-(2.36). Eq. (2.19) implies that to derive the interface conditions of $Z_{\mathrm{p,s}}$ will be equivalent to derive the interface conditions for $\partial_{\boldsymbol{z}} \boldsymbol{Q}_{\mathrm{p,s}}$ and $\partial_{\boldsymbol{z}} \boldsymbol{P}_{\mathrm{p,s}}$, which requires to use the conservation of level set functions designed in the Eulerian frozen Gaussian approximation formula [60, 97].

3.2.1 Interface conditions for position and momentum

For simplification of the notation we only work with the plus branch and $\mathbf{P}^p = \mathbf{P}$, $\mathbf{P}^v = \mathbf{S}^v$, $\mathbf{P}^h = \mathbf{S}^h$. Thus for each wavefield at the interface, The cases to consider are pre-critical, critical and post-critical. For each incident wavefield we consider all cases. For the following we only focus on an incoming P wave as an incoming SV wave is similar in derivation. To simplify notation for this section we let $\mathbf{S}^{v,m} = \mathbf{S}^m$. An application of Snell's law gives; Let us explain the formulation of reflection and transmission interface

conditions with more details for the incident P-wave as illustrated in Fig. 3.1. In this case, if one denotes $(\mathbf{Q}_{\mathrm{p,s}}^{\mathrm{in,re,tr}}, \mathbf{P}_{\mathrm{p,s}}^{\mathrm{in,re,tr}}, a_{\mathrm{p,s}}^{\mathrm{in,re,tr}})$ to be the center, propagation vector and prefactor amplitude of the corresponding incident, reflected and transmitted Gaussian wave packets for P- and S-waves respectively, then $\mathbf{Q}^{\mathrm{in}} = \mathbf{Q}^{\mathrm{re}} = \mathbf{Q}^{\mathrm{tr}}$ and $\mathbf{P}^{\mathrm{re,tr}}$ is updated as follows

$$\mathbf{P}_{\mathrm{p,s}}^{\mathrm{tr,re}} = \mathbf{P}^{\mathrm{in}} + \mathrm{sgn}(\mathbf{P}_{\mathrm{p,s}}^{\mathrm{tr,re}})\Big(\sqrt{|\mathbf{P}^{\mathrm{in}}|n_{\mathrm{p,s}}^{\mathrm{tr,re}} - \big||\mathbf{P}^{\mathrm{in}}| - (\mathbf{P}^{\mathrm{in}} \cdot \mathbf{N})^{\mathbf{2}}\big|} - (\mathbf{P} \cdot \mathbf{N})\Big)\mathbf{N} \tag{3.2}$$

where $n_{\mathrm{p,s}}^{\mathrm{tr,re}}$ denotes the index of refraction for the new respective direction, e.g. $n_{\mathrm{p}}^{\mathrm{tr}} = c_{\mathrm{p}}^{\vee}/c_{\mathrm{p}}^{\wedge}$.

3.2.2 Interface condition for the amplitude

Let $\theta_i, \theta_r, \theta_t$ denote the P wave incident, reflection and transmission angle respectively and ϕ_r, ϕ_t denote the SV wave reflection and transmission angle respectively. There are two critical angles that can occur. These give rise to $\mathbf{P}^t$ traveling along the interface instead of being transmitted and $\mathbf{S}^t$ traveling along the interface instead of being transmitted. Denote these angles as β_1, β_2 respectively. In all cases we have

$$\eta_2^p = (c_{\mathrm{p}}^{\vee} c_{\mathrm{p}}^{\wedge})^{-2}\sqrt{(c_{\mathrm{p}}^{\vee})^2 - (c_{\mathrm{p}}^{\wedge})^2 \sin^2(\theta_i)} \tag{3.3}$$

$$\eta_2^s = (c_{\mathrm{p}}^{\vee} c_{\mathrm{s}}^{\wedge})^{-2}\sqrt{(c_{\mathrm{p}}^{\vee})^2 - (c_{\mathrm{s}}^{\wedge})^2 \sin^2(\theta_i)} \tag{3.4}$$

For the pre-critical , $\theta_i < \beta_1 < \beta_2$. For the first critical case, we see

$$\theta_i = \beta_1 < \beta_2 \tag{3.5}$$

so $\eta_2^p = 0$ and the second critical case

$$\beta_1 < \beta_2 = \theta_i \tag{3.6}$$

so $\eta_2^p = 0$. If η_2^p or η_2^s is imaginary there is no transmitted P/S wave respectively. Using the potentials, let $\Phi = (\Phi_1, \Phi_2)$, $\Psi = (\Psi_1, \Psi_2)$ in (x,z)-coordinates. The potentials must satisfy the equations;

$$c_{\mathrm{p}}^{\vee}(\partial_x^2\Phi + \partial_z^2\Phi) = \partial_t^2\Phi \tag{3.7}$$

$$c_{\mathrm{s}}^{\vee}(\partial_x^2\Psi + \partial_z^2\Psi) = \partial_t^2\Psi \tag{3.8}$$

So the plane wave solution is given component-wise by

$$\Phi_1 = a_p^i \exp\Big(\mathrm{i}/\epsilon\big(\sin(\theta_i)x + \cos(\theta_i)z - c_{\mathrm{p}}^{\vee}t\big)\Big) \tag{3.9}$$

$$+\ a_p^r \exp\Big(\mathrm{i}/\epsilon\big(\sin(\theta_r)x - \cos(\theta_r)z - c_{\mathrm{p}}^{\vee}t\big)\Big) \tag{3.10}$$

$$\Phi_2 = a_p^t \exp\Big(\mathrm{i}/\epsilon\big(\sin(\theta_t)x + \cos(\theta_t)z - c_{\mathrm{p}}^{\wedge}t\big)\Big) \tag{3.11}$$

$$\Psi_1 = a_s^r \exp\Big(\mathrm{i}/\epsilon\big(\sin(\phi_r)x - \cos(\phi_r)z - c_{\mathrm{s}}^{\vee}t\big)\Big) \tag{3.12}$$

$$\Psi_2 = a_s^t \exp\Big(\mathrm{i}/\epsilon\big(\sin(\phi_t)x + \cos(\phi_t)z - c_{\mathrm{s}}^{\wedge}t\big)\Big) \tag{3.13}$$

At the interface we have the displacement is continuous

$$\partial_x\Phi_1 + \partial_z\Psi_1 = \partial_x\Phi_2 + \partial_z\Psi_2 \tag{3.14}$$

$$\partial_z\Phi_1 - \partial_x\Psi_1 = \partial_z\Phi_2 - \partial_x\Psi_2 \tag{3.15}$$

With Snells law, this gives;

$$\frac{\sin(\theta_r)}{c_{p,1}}a_p^i + \frac{\sin(\theta_r)}{c_{p,1}}a_p^r - \frac{\cos(\phi_r)}{c_{s,1}}a_s^r = \frac{\sin(\theta_t)}{c_{p,2}}a_p^t + \frac{\cos(\phi_t)}{c_{s,2}}a_s^t \tag{3.16}$$

$$\frac{\cos(\theta_r)}{c_{p,1}}a_p^i - \frac{\cos(\theta_r)}{c_{p,1}}a_p^r - \frac{\sin(\phi_r)}{c_{s,1}}a_s^r = \frac{\cos(\theta_t)}{c_{p,2}}a_p^t - \frac{\sin(\phi_t)}{c_{s,2}}a_s^t \tag{3.17}$$

The normal and tangential stress components are continuous, looking back at the equation with Hooke's law

$$\lambda\Delta + 2\mu\epsilon_{zz} = (\lambda + 2\mu)\left(\partial_z^2\Phi + \lambda\partial_x^2\Psi\right) - 2\mu\partial_{xz}^2\Psi \tag{3.18}$$

$$2\mu\epsilon_{xz} = \mu\left(2\partial_{xz}^2\Phi + \partial_z^2\Psi - \partial_x^2\Psi\right) \tag{3.19}$$

Enforcing continuity component wise gives;

$$\rho_1\partial_t^2\Phi_1 - 2\rho_1 c_{s,1}^2\left(\partial_x^2\Phi_1 + \partial_{xz}^2\Psi_1\right) = \rho_2\partial_t^2\Phi_2 - 2\rho_2 c_{s,2}^2\left(\partial_x^2\Phi_2 + \partial_{xz}^2\Psi_2\right) \tag{3.20}$$

$$\mu_1\left(2\partial_{xz}^2\Phi_1 + \partial_z^2\Psi_1 - \partial_x^2\Psi_1\right) = \mu_2\left(2\partial_{xz}^2\Phi_2 + \partial_z^2\Psi_2 - \partial_x^2\Psi_2\right) \tag{3.21}$$

plugging in the potentials and simplifying

$$-\cos(2\phi_r)a_p^i - \cos(2\phi_r)a_r^p - \sin(2\phi_r)a_r^s = -\frac{\rho_2}{\rho_1}\cos(2\phi_t)a_t^p + \frac{\rho_2}{\rho_1}\sin(2\phi_t)a_t^s \tag{3.22}$$

$$-\sin(2\theta_r)a_p^i + \sin(2\theta_r)a_r^p - \frac{c_{p,1}^2}{c_{s,1}^2}\cos(2\phi_r)a_r^s = \frac{\rho_2 c_{s,2}^2 c_{p,1}^2}{\rho_1 c_{p,2}^2 c_{s,1}^2}\sin(2\theta_t)a_t^p - \frac{\rho_2 c_{p,1}^2}{\rho_1 c_{s,1}^2}\cos(2\phi_t)a_t^s \tag{3.23}$$

which gives the Zoeppritz equations

$$M \begin{pmatrix} a_p^r \\ a_s^r \\ a_p^t \\ a_s^t \end{pmatrix} = \begin{pmatrix} \cos(\theta_r) \\ \sin(\theta_r) \\ \cos(2\phi_r) \\ \cos(2\theta_r) \end{pmatrix} a_p^i \tag{3.24}$$

Where the matrix M is

$$\begin{pmatrix} \cos(\theta_r) & \frac{c_{\mathrm{p}}^{\vee}}{c_{\mathrm{s}}^{\vee}} \sin(\phi_r) & \frac{c_{\mathrm{p}}^{\vee}}{c_{\mathrm{p}}^{\wedge}} \cos(\theta_t) & -\frac{c_{\mathrm{p}}^{\vee}}{c_{\mathrm{s}}^{\wedge}} \sin(\phi_t) \\ -\sin(\theta_r) & \frac{c_{\mathrm{p}}^{\vee}}{c_{\mathrm{s}}^{\vee}} \cos(\phi_r) & \frac{c_{\mathrm{p}}^{\vee}}{c_{\mathrm{p}}^{\wedge}} \sin(\theta_t) & \frac{c_{\mathrm{p}}^{\vee}}{c_{\mathrm{s}}^{\wedge}} \sin(\phi_r) \\ -\cos(2\phi_r) & -\sin(2\phi_r) & \frac{\rho_2}{\rho_1} \cos(2\phi_t) & -\frac{\rho_2}{\rho_1} \sin(2\phi_t) \\ \sin(2\theta_r) & -(\frac{c_{\mathrm{p}}^{\vee}}{c_{\mathrm{s}}^{\vee}})^2 \cos(2\phi_r) & \frac{\rho_2 (c_{\mathrm{s}}^{\wedge} c_{\mathrm{p}}^{\vee})^2}{\rho_1 (c_{\mathrm{p}}^{\wedge} c_{\mathrm{s}}^{\vee})^2} \sin(2\theta_t) & \frac{\rho_2 (c_{\mathrm{p}}^{\vee})^2}{\rho_2 (c_{\mathrm{s}}^{\vee})^2} \sin(\phi_t) \end{pmatrix} \tag{3.25}$$

The reflection and transmission coefficients are given by

$$\begin{pmatrix} R_p \\ R_s \\ T_p \\ T_s \end{pmatrix} = M^{-1} \begin{pmatrix} \cos(\theta_r) \\ \sin(\theta_r) \\ \cos(2\phi_r) \\ \cos(2\theta_r) \end{pmatrix} \tag{3.26}$$

So

$$\begin{pmatrix} a_p^r \\ a_s^r \\ a_p^t \\ a_s^t \end{pmatrix} = \begin{pmatrix} R_p \\ R_s \\ T_p \\ T_s \end{pmatrix} a_p^i \tag{3.27}$$

This is of course provided $\theta_i < \beta_1$. Otherwise the coefficients become much simper, e.g. if $\beta_1 < \theta_i < \beta_2$, R_s, R_t remain the same while $R_p = -1$ and $T_p = 0$.

3.2.3 Derivation of the interface conditions for $\partial_{\boldsymbol{z}}\boldsymbol{Q}$ and $\partial_{\boldsymbol{z}}\boldsymbol{P}$

The interface conditions for $\partial_{\boldsymbol{z}}\boldsymbol{Q}_{\mathrm{p,s}}$ and $\partial_{\boldsymbol{z}}\boldsymbol{P}_{\mathrm{p,s}}$ are as follows:

$$\begin{aligned}
&\partial_{\mathbf{z}}\mathbf{Q}^{\mathrm{re,tr}} = \partial_{\mathbf{z}}\mathbf{Q}^{\mathrm{in}}\, F,\\
&\partial_{\mathbf{z}}\mathbf{P}^{\mathrm{re,tr}} = \partial_{\mathbf{z}}\mathbf{P}^{\mathrm{in}}\, W - \frac{|\mathbf{P}^{\mathrm{re,tr}}|}{c(\mathbf{Q}^{\mathrm{re,tr}})p_z^{\mathrm{re,tr}}}\left(\partial_{\mathbf{z}}\mathbf{Q}^{\mathrm{re,tr}}\cdot\nabla c(\mathbf{Q}^{\mathrm{re,tr}}) - \partial_{\mathbf{z}}\mathbf{Q}^{\mathrm{in}}\cdot\nabla c(\mathbf{Q}^{\mathrm{in}})\right)\boldsymbol{e}_3,
\end{aligned} \tag{3.28}$$

where $(\mathbf{Q}^{\mathrm{in,re,tr}}, \mathbf{P}^{\mathrm{in,re,tr}})$ corresponds to the center and propagation vector of incident, reflected and transmitted Gaussian wave packet for either P- or S-waves, respectively, and F, W are two 3×3 matrices with $F^T = W^{-1}$, and

$$F = \begin{bmatrix} & 1 & \\ & & 1 \\ (\kappa-1)\frac{p_x}{p_z^{\mathrm{in}}} & (\kappa-1)\frac{p_y}{p_z^{\mathrm{in}}} & \kappa\frac{p_z^{\mathrm{re,tr}}}{p_z^{\mathrm{in}}} \end{bmatrix}, \quad \text{with} \quad \kappa = \left(\frac{c(\mathbf{Q}^{\mathrm{re,tr}})}{c(\mathbf{Q}^{\mathrm{in}})}\right)^2.$$

The derivation requires to use the conservation of level set functions designed in the Eulerian frozen Gaussian approximation formula [60, 97], where the idea is to use the following Liouville operator to describe the dynamics of Gaussian wave packet on phase plane,

$$\mathcal{L}_{\mathrm{p,s}} = \partial_t + \partial_{\boldsymbol{p}}H_{\mathrm{p,s}}\cdot\partial_{\boldsymbol{q}} - \partial_{\boldsymbol{q}}H_{\mathrm{p,s}}\cdot\partial_{\boldsymbol{p}} \tag{3.29}$$

whose corresponding characteristic equations are given by the Hamiltonian systems eqs. (2.14) with $H_{\mathrm{p,s}} = \pm c_{\mathrm{p,s}}(\boldsymbol{q})|\boldsymbol{p}|$. For example, the prefactor equation of the P-wave for the "+"

wave propagation direction is given by, in the Eulerian formulation,

$$\mathcal{L}a_{\mathrm{p}} = a_{\mathrm{p}}\left(\frac{\partial_{\mathbf{Q}_{\mathrm{p}}} c_{\mathrm{p}}(\mathbf{Q}_{\mathrm{p}})\cdot \mathbf{P}_{\mathrm{p}}}{|\mathbf{P}_{\mathrm{p}}|} + \frac{1}{2}\,\mathrm{tr}\left(Z_{\mathrm{p}}^{-1}\partial_t Z_{\mathrm{p}}\right)\right).$$

We shall derive the interface conditions of $\partial_{\boldsymbol{z}}\boldsymbol{Q}$ and $\partial_{\boldsymbol{z}}\boldsymbol{P}$ for the transmitted P-wave and the "+" wave propagation direction, and omit the subscript "p" in the following derivation for a sake of simplicity. Consider a level set function $\boldsymbol{\phi}(t,\mathbf{q},\mathbf{p}) = (\phi_1,\phi_2,\phi_3)$ which satisfies

$$\mathcal{L}\boldsymbol{\phi} = 0, \quad \text{with} \quad \boldsymbol{\phi}(0,\mathbf{q},\mathbf{p}) = \mathbf{p} + \mathrm{i}\mathbf{q}, \tag{3.30}$$

then the Eulerian formulation of FGA [60] shows that

$$\partial_{\mathbf{z}}\mathbf{Q} = (\partial_{\mathbf{p}}\boldsymbol{\phi})^{\mathrm{T}}, \qquad \partial_{\mathbf{z}}\mathbf{P} = -(\partial_{\mathbf{q}}\boldsymbol{\phi})^{\mathrm{T}}. \tag{3.31}$$

We will follow the strategy described in [12, 97], and consider the case illustrated in Fig. 3.1, where the level set functions $\boldsymbol{\phi}^{\mathrm{re,tr}}$ for the transmitted P-waves satisfy the same evolution as $\boldsymbol{\phi}$ in eq. (3.30) with the following interface conditions

$$\boldsymbol{\phi}^{\mathrm{tr}}(t,\mathbf{q}^{\mathrm{tr}},\mathbf{p}^{\mathrm{tr}}) = \boldsymbol{\phi}^{\mathrm{tr}}(t,\mathbf{q}^{\mathrm{in}},\mathbf{p}^{\mathrm{in}}). \tag{3.32}$$

Differentiating eq. (3.32) by the definition of partial derivatives and chain rule, and making use of eq. (3.2) yield

$$\begin{aligned}
\partial_{p_x}\boldsymbol{\phi}^{\mathrm{tr}}(t,\mathbf{q}^{\mathrm{tr}},\mathbf{p}^{\mathrm{tr}}) &= \left(n_1^{-2}-1\right)\frac{p_x}{p_z}\partial_{p_z}\boldsymbol{\phi}^{\mathrm{tr}}(t,\mathbf{q}^{\mathrm{in}},\mathbf{p}^{\mathrm{in}}) + \partial_{p_x}\boldsymbol{\phi}^{\mathrm{tr}}(t,\mathbf{q}^{\mathrm{in}},\mathbf{p}^{\mathrm{in}}),\\
\partial_{p_y}\boldsymbol{\phi}^{\mathrm{tr}}(t,\mathbf{q}^{\mathrm{tr}},\mathbf{p}^{\mathrm{tr}}) &= \left(n_1^{-2}-1\right)\frac{p_y}{p_z}\partial_{p_z}\boldsymbol{\phi}^{\mathrm{tr}}(t,\mathbf{q}^{\mathrm{in}},\mathbf{p}^{\mathrm{in}}) + \partial_{p_y}\boldsymbol{\phi}^{\mathrm{tr}}(t,\mathbf{q}^{\mathrm{in}},\mathbf{p}^{\mathrm{in}}),\\
\partial_{p_z}\boldsymbol{\phi}^{\mathrm{tr}}(t,\mathbf{q}^{\mathrm{tr}},\mathbf{p}^{\mathrm{tr}}) &= n_1^{-2}\frac{p_z^{\mathrm{tr}}}{p_z}\partial_{p_z}\boldsymbol{\phi}^{\mathrm{tr}}(t,\mathbf{q}^{\mathrm{in}},\mathbf{p}^{\mathrm{in}}).
\end{aligned} \tag{3.33}$$

Moreover, differentiating eq. (3.32) with respect to t gives

$$\partial_t \boldsymbol{\phi}^{\mathrm{tr}}(t, \mathbf{q}^{\mathrm{tr}}, \mathbf{p}^{\mathrm{tr}}) = \partial_t \boldsymbol{\phi}^{\mathrm{tr}}(t, \mathbf{q}^{\mathrm{in}}, \mathbf{p}^{\mathrm{in}}), \tag{3.34}$$

which implies by eq. (3.29), at the interface,

$$\left[\nabla_{\mathbf{p}} H \cdot \nabla_{\mathbf{q}} \boldsymbol{\phi}^{\mathrm{tr}}\right]_{q_z = z_0} = \left[\nabla_{\mathbf{q}} H \cdot \nabla_{\mathbf{p}} \boldsymbol{\phi}^{\mathrm{tr}}\right]_{q_z = z_0}, \tag{3.35}$$

with $[\cdot]$ denoting the jump function. Therefore,

$$\begin{aligned}
\partial_{q_x} \boldsymbol{\phi}^{\mathrm{tr}}(t, \mathbf{q}^{\mathrm{tr}}, \mathbf{p}^{\mathrm{tr}}) &= \partial_{q_x} \boldsymbol{\phi}^{t}(t, \mathbf{q}^{\mathrm{in}}, \mathbf{p}^{\mathrm{in}}), \\
\partial_{q_y} \boldsymbol{\phi}^{\mathrm{tr}}(t, \mathbf{q}^{\mathrm{tr}}, \mathbf{p}^{\mathrm{tr}}) &= \partial_{q_y} \boldsymbol{\phi}^{t}(t, \mathbf{q}^{\mathrm{in}}, \mathbf{p}^{\mathrm{in}}), \\
\partial_{q_z} \boldsymbol{\phi}^{\mathrm{tr}}(t, \mathbf{q}^{\mathrm{tr}}, \mathbf{p}^{\mathrm{tr}}) &= n_1^2 \frac{p_z}{p_z^{\mathrm{tr}}} \partial_{q_z} \boldsymbol{\phi}^{\mathrm{tr}}(t, \mathbf{q}^{\mathrm{in}}, \mathbf{p}^{\mathrm{in}}) + (n_1^2 - 1) \left(\frac{p_x}{p_z^{\mathrm{tr}}} \partial_{q_x} + \frac{p_y}{p_z^{\mathrm{tr}}} \partial_{q_y} \right) \boldsymbol{\phi}^{\mathrm{tr}}(t, \mathbf{q}^{\mathrm{in}}, \mathbf{p}^{\mathrm{in}}) \\
&\quad + \frac{|\mathbf{p}^{\mathrm{tr}}|}{c^{\vee} p_z^{\mathrm{tr}}} \left[|\mathbf{p}| \nabla_{\mathbf{q}} c(\mathbf{q}) \cdot \nabla_{\mathbf{p}} \boldsymbol{\phi}^{\mathrm{tr}}(t, \mathbf{q}, \mathbf{p}) \right]_{q_z = z_0}.
\end{aligned} \tag{3.36}$$

Then eqs. (3.31) and (3.33) imply the interface condition for $\partial_z \mathbf{Q}$ in eq. (3.28), while eqs. (3.31) and (3.36) imply the interface condition for $\partial_z \mathbf{P}$ in eq. (3.28) for the transmitted P-wave. The other cases including the interface conditions for the reflected P-wave, transmitted and reflected S-wave can be derived in an essentially same way.

3.3 Greens Function from a point source

In this section we derive greens function from a point source. The solution derived (3.48) will give and equation to be used as the initial displacement $\mathbf{u}_0^{\epsilon}$ and velocity $\mathbf{u}_1^{\epsilon}$ that will propagate outward form the initial wavefield. For simplicity, consider the

elastic wave equation with a forcing term $\mathbf{F}(t)\delta(x_0)$ with $\mathbf{F}(t)$ and constant density ρ.

$$\rho\partial_{tt}u = (\lambda + 2\mu)\nabla(\nabla\cdot\mathbf{u}) - \mu\,\nabla\times\nabla\times\mathbf{u} + F(t)\delta_{x_0} \tag{3.37}$$

Let $\Delta\Phi = \nabla\cdot u$, $\nabla\times\Psi = \nabla\times u$, and define the orthogonal operators

$$P = \nabla(\nabla\cdot), \qquad S = \Delta - \nabla(\nabla\cdot) = -\nabla\times\nabla\times \tag{3.38}$$

Then eq (3.37) can be written in terms of the operators (3.38), which gives

$$\rho\partial_{tt}u = (\lambda + 2\mu)P\mathbf{u} + \mu S\mathbf{u} + \mathbf{F}(t)\delta_{x_0} \tag{3.39}$$

Assuming the solution is $\mathbf{u}(t,\mathbf{x}) = \nabla\Phi + \Psi$, with $\nabla\cdot\Psi = 0$. Then

$$P\mathbf{u} = \nabla(\Delta\Phi), \qquad S\mathbf{u} = -\nabla\times\nabla\times\Psi = \Delta\Psi \tag{3.40}$$

Writing $\partial_{tt}\mathbf{u} = \nabla\partial_{tt}\Phi + \partial_{tt}\Psi$ and plugging in gives the equation

$$\rho(\nabla\partial_{tt}\Phi + \partial_{tt}\Psi) = (\lambda + 2\mu)\nabla(\Delta\Phi) + \mu\Delta\Psi + \mathbf{F}(t)\delta_{x_0} \tag{3.41}$$

Then eq (3.37) can then be written as,

$$\rho(c_p^2\Delta - \partial_{tt})\nabla\Phi + \rho(c_s^2\Delta - \partial_{tt})\Psi = \mathbf{F}(t)\delta_{x_0}. \tag{3.42}$$

With the previous notation of $\mathbf{u} = P\mathbf{a}_p + S\mathbf{a}_s$, this becomes

$$\rho P(c_{\mathrm{p}}^2\Delta - \partial_{tt})\mathbf{a}_{\mathrm{p}} + \rho S(c_{\mathrm{s}}^2\Delta - \partial_{tt})\mathbf{a}_{\mathrm{s}} = F(t)\delta_{x_0}, \tag{3.43}$$

so $P\mathbf{a}_{\mathrm{p}} = \nabla\Phi$ and $S\mathbf{a}_{\mathrm{s}} = \Psi$. Writing the forcing term in terms of the Laplacian,

$$\mathbf{F}(t)\delta_{x_0} = -\Delta\left(\frac{\mathbf{F}(t)}{4\pi r}\right) = -(P+S)\left(\frac{\mathbf{F}(t)}{4\pi r}\right) \tag{3.44}$$

with $r = |\mathbf{x} - \mathbf{x}_0|$. Thus, there are two vector wave equations to solve. Looking for radial symmetry, and writing the Laplacian in polar

$$(c^2\partial_{rr} - \partial_{tt})(ra) = \frac{\mathbf{F}}{4\pi\rho} \tag{3.45}$$

The general solution is written as

$$r\mathbf{a} = -\frac{\mathbf{G}(t)}{4\pi r\rho} + \tilde{\phi}(t - r/c) + \tilde{\psi}(t + r/c) \tag{3.46}$$

with $\partial_{tt}\mathbf{G}(t) = \mathbf{F}(t)$ and $\mathbf{F}(0) = \partial_t\mathbf{F}(0) = 0$. Assuming no incoming waves let $\tilde{\psi} = 0$, letting $r \to 0$ leads to $\tilde{\phi} = \mathbf{G}/4\pi r\rho$. So

$$\mathbf{a}_{\mathrm{p}} = \frac{\mathbf{G}(t - r/c_p) - \mathbf{G}(t)}{4\pi\rho r}, \qquad \mathbf{a}_s = \frac{\mathbf{G}(t - r/c_s) - \mathbf{G}(t)}{4\pi\rho r} \tag{3.47}$$

As $\mathbf{u} = P\mathbf{a}_{\mathrm{p}} + S\mathbf{a}_{\mathrm{s}}$, differentiating and writing in the component form;

$$u_i = \frac{x_i x_j}{4\pi\rho c_p^2 r^3}F_j(t - r/c_p) + \frac{r^2\delta_{ij} - x_i x_j}{4\pi\rho c_s^2 r^3}F_j(t - r/c_s) + \frac{3x_i x_j - r^2\delta_{ij}}{4\pi\rho c_s^2 r^5}\int_{r/c_p}^{r/c_s} sF_j(t - s)\,\mathrm{d}s. \tag{3.48}$$

3.4 Seismic Inversion

In this section full waveform inversion (FWI) and traveltime tomography are summarized, see [24] for complete details.

3.4.1 Full Waveform Inversion

As it makes the notation more readable we do the summary with the equation (2.64) in its more general form.

$$\rho(\mathbf{x})\partial_t^2\mathbf{u}(\mathbf{x},t) - \nabla\cdot\boldsymbol{\sigma}(\mathbf{x},t) = \mathbf{F}(\mathbf{x},t) \tag{3.49}$$

with the stress tensor

$$\boldsymbol{\sigma}(\mathbf{x},t) = \int_{t_0}^{\infty} \partial_t\mathbf{C}(\mathbf{x},t-\tau) : \nabla(\mathbf{x},\tau)\ \mathrm{d}\tau \tag{3.50}$$

where $\mathbf{C}$ is the elasticity tensor. Equation (3.49) can now be written as $\mathbf{L}(\mathbf{u},\rho,\mathbf{C}) = \mathbf{F}$ with

$$\mathbf{L}(\mathbf{u},\rho,\mathbf{C}) = \rho(\mathbf{x})\partial_t^2\mathbf{u}(\mathbf{x},t) - \nabla\cdot\int_{t_0}^{\infty} \partial_t\mathbf{C}(\mathbf{x},t-\tau) : \nabla\mathbf{u}(\mathbf{x},\tau)\ d\tau \tag{3.51}$$

Denote $\delta\mathbf{d} = \mathbf{u}_{obs} - \mathbf{u}_{sys}$, with $\mathbf{u}_{obs}(m;\mathbf{x},t)$ being data observed from the seismogram, $\mathbf{u}_{sys}(m;\mathbf{x},t)$ the data from a synthetic seismogram at the receiver located at $\mathbf{x}_r$ and m denoting the model parameters; such as, ρ,λ,μ. Let Ω represent the computational

domain. The misfit functional

$$\begin{aligned}\chi(m) = \chi(\mathbf{u}(m)) &= \frac{1}{2}\langle \delta\mathbf{d}, \delta\mathbf{d}\rangle \\ &= \frac{1}{2}\int_\Omega \int_0^T (\mathbf{u}_{obs} - \mathbf{u}_{sys})^2 \delta(\mathbf{x} - \mathbf{x}_r)\ \mathrm{d}t\,\mathrm{d}\mathbf{x} \\ &= \int_\Omega \int_0^T \mathbf{s}(m)\ \mathrm{d}t\,\mathrm{d}\mathbf{x}\end{aligned} \tag{3.52}$$

With spacial integration defined on the data space Ω. The object of FWI is to minimize the misfit functional. Denote $\nabla_m \chi \delta m$ as the derivative of $\chi(m)$ with respect to m in the direction of δm. Taking ∇_m of equation (3.49) and using the adjoint operators and linearity we can arrive at the adjoint equation

$$\rho\mathbf{u}^\dagger - \nabla\cdot\boldsymbol{\sigma}^\dagger = -\nabla^\dagger_{\mathbf{u}}\mathbf{s}^\dagger(m) \tag{3.53}$$

with $\mathbf{u}^\dagger$ the adjoint wavefield, the adjoint stress tensor defined as

$$\sigma^\dagger = \int_t^T \partial_t \mathbf{C}(\tau - t) : \nabla\mathbf{u}^\dagger(\tau)\ \mathrm{d}\tau \tag{3.54}$$

and $\mathbf{s}^\dagger(m)$ the adjoint source. With the solution of the adjoint equation, $\mathbf{u}^\dagger$ the derivative of the misfit kernel is cast as;

$$\nabla_m \chi \delta m = \int_\Omega \int_0^T \mathbf{u}^\dagger \cdot \nabla_m \mathbf{L}\delta m\ \mathrm{d}t\,\mathrm{d}\mathbf{x} \tag{3.55}$$

In terms of the sensitivity kernels, this is

$$\nabla_m \chi \delta m = \int_\Omega K_m(\mathbf{x})\delta m(\mathbf{x})\ \mathrm{d}\mathbf{x} \tag{3.56}$$

with

$$K_m = \int_0^T \mathbf{u}^\dagger \cdot \nabla_m \mathbf{L} \delta m \ \mathrm{d}t \tag{3.57}$$

For an isotropy the elasticity tensor becomes

$$\mathbf{C} = C_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu \delta_{ik} \delta_{jl} + \mu \delta_{il} \delta_{jk} \tag{3.58}$$

In general, ρ, λ, μ can be thought of as independent parameters, i.e.

$$\nabla_{\mathbf{m}} \chi \delta \mathbf{m} = \nabla_\rho \chi \delta \rho + \nabla_\lambda \chi \delta \lambda + \nabla_\mu \chi \delta \mu \tag{3.59}$$

with

$$\nabla_\rho \chi \delta \rho = - \int_\Omega \int_0^T \delta \rho \partial_t \mathbf{u}^\dagger \cdot \partial_t \mathbf{u} \ \mathrm{d}t \, \mathrm{d}\mathbf{x}, \tag{3.60}$$

$$\nabla_\lambda \chi \delta \lambda = \int_\Omega \int_0^T \delta \lambda (\nabla \cdot \mathbf{u}^\dagger)(\nabla \cdot \mathbf{u}) \ \mathrm{d}t \, \mathrm{d}\mathbf{x}, \tag{3.61}$$

$$\nabla_\mu \chi \delta \mu = \int_\Omega \int_0^T \delta \mu \big[(\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u}) + (\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u})^T \big] \ \mathrm{d}t d\mathbf{x}. \tag{3.62}$$

This gives sensitivity kernels

$$K_\rho = - \int_0^T \partial_t \mathbf{u}^\dagger \cdot \partial_t \mathbf{u} \ \mathrm{d}t, \tag{3.63}$$

$$K_\lambda = \int_0^T (\nabla \cdot \mathbf{u}^\dagger)(\nabla \cdot \mathbf{u}) \ \mathrm{d}t, \tag{3.64}$$

$$K_\mu = \int_0^T \big[(\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u}) + (\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u})^T \big] \ \mathrm{d}t. \tag{3.65}$$

With a given iterative method, the misfit kernel is used to update the search direction. For simplicity we assume λ and μ are constant so only the misfit kernel $\nabla_\rho \chi \delta \rho$ is needed.

The Born approximation can be used to minimize the misfit functional, Let ρ_0 be an initial guess, then

$$\chi(\rho_0 + \delta\rho) \approx \chi(\rho) + \nabla_\rho \chi(\rho_0) \cdot \delta\rho + \frac{1}{2}\delta\rho \cdot \nabla^2_\rho \chi(\rho_0)\delta\rho \tag{3.66}$$

Considering the expansion for the $\nabla_\rho \chi$ at a minimum, the update for ρ is given by

$$\delta\rho = -\left(\nabla^2_\rho \chi(\rho_0)\right)^{-1} \nabla_\rho \chi(\rho_0) \tag{3.67}$$

and the gradient and Hessian can be computed by

$$\nabla_\rho \chi = -\left\langle \frac{\delta \mathbf{u}_{sys}}{\delta\rho}, \delta\boldsymbol{d} \right\rangle \tag{3.68}$$

$$\nabla^2_\rho \chi = \left\langle \frac{\delta \mathbf{u}_{sys}}{\delta\rho}, \frac{\delta \mathbf{u}_{sys}}{\delta\rho} \right\rangle - \left\langle \frac{\delta^2 \mathbf{u}_{sys}}{\delta^2\rho}, \boldsymbol{d} \right\rangle \tag{3.69}$$

For the linear perturbation model, the second term in eq (3.69) is zero, this gives the update;

$$\delta\rho = -\left(\left\langle \frac{\delta \mathbf{u}_{sys}}{\delta\rho}, \frac{\delta \mathbf{u}_{sys}}{\delta\rho} \right\rangle\right)^{-1} \left\langle \frac{\delta \mathbf{u}_{sys}}{\delta\rho}, \delta\boldsymbol{d} \right\rangle \tag{3.70}$$

3.4.2 Traveltime adjoint tomography

For traveltime adjoint tomography, the only the traveltime information is used, similar kernels can be defined using,

$$\mathbf{u}_{obs/sys} = T_{obs/sys} \qquad \delta\boldsymbol{d} = T_{obs} - T_{sys} \qquad \langle \delta\boldsymbol{d}, \delta\boldsymbol{d} \rangle = |\delta T|^2 \tag{3.71}$$

The forward model can than be solved (3.37) with zero initial conditions; i.e.,

$$\begin{cases} (\rho(\mathbf{x})\partial_t^2 - \mathcal{L})\mathbf{u}(t,\mathbf{x}) = \mathbf{F}(t)\delta_d(\mathbf{x}-\mathbf{x}_s) \\ \mathbf{u}(0,\mathbf{x}) = \partial_t\mathbf{u}(0,\mathbf{x}) = 0 \end{cases} \tag{3.72}$$

with as the $\mathbf{x}_s$ source location. Then the adjoint equation, which represents backwards propagation, can be solved with zero initial conditions; i.e.,

$$\begin{cases} (\rho(\mathbf{x})\partial_t^2 - \mathcal{L})\mathbf{u}^\dagger(t,\mathbf{x}) = \mathbf{s}^\dagger(t)\delta_d(\mathbf{x}-\mathbf{x}_r) \\ \mathbf{u}^\dagger(T,\mathbf{x}) = \partial_t\mathbf{u}^\dagger(T,\mathbf{x}) = 0 \end{cases} \tag{3.73}$$

where $\mathbf{x}_r$ is the receiver location, the adjoint source $\mathbf{s}^\dagger(t) = w(t)\partial_t u(t,\mathbf{x})/\mathcal{W}_r$, with $w(t)$ as a weight function on $[0,T]$ and

$$\mathcal{W}_r = \int_0^T w(t)\mathbf{u}(t,\mathbf{x}_r)\partial_t^2\mathbf{u}(t,\mathbf{x}_r)\,\mathrm{d}t. \tag{3.74}$$

The sensitivity kernel can then be computed in a similar fashion as the previous section, and is

$$K_\rho = -\int_0^T 2\rho(\mathbf{x})\partial_t\mathbf{u}^\dagger(t,\mathbf{x})\cdot\partial_t\mathbf{u}(t,\mathbf{x})\,\mathrm{d}t. \tag{3.75}$$

The update can be given by the least squares solution using

$$\delta\boldsymbol{d} = \int_\Omega K_\rho\delta\rho\,\mathrm{d}\mathbf{x}. \tag{3.76}$$

Chapter 4

Numerical Experiments

4.1 Techniques for Implementation

Several components are need to construct the FGA, first a mesh in the phase space $(\mathbf{q}, \mathbf{p})$. With careful consideration of the initial conditions this can be done in an efficient manner. For localized initial conditions $\mathbf{u}_0^\epsilon$, $\mathbf{u}_1^\epsilon$ define an appropriate $\mathbf{q}$-mesh $\{\mathbf{q}^{\mathbf{j}_k}\}$ for $\mathbf{j} = (j_1, j_2, j_3)$ and $\mathbf{j}_k = 1$ to Nq. If WKBJ initial wavefields are given known; that is,

$$\mathbf{u}_0^\epsilon(\mathbf{x}) = A_p(\mathbf{x}) \exp(\mathrm{i}/\epsilon \Phi_p(\mathbf{x})) + A_s(\mathbf{x}) \exp(\mathrm{i}/\epsilon \Phi_s(\mathbf{x})) \tag{4.1}$$

one can localize the p-mesh around the gradient of the phase functions.

$$\mathbf{p}^{\mathbf{j},\mathbf{l}} = \{\nabla \Phi_p(\mathbf{q}_\mathbf{l}) + \mathbf{l}\delta p\} \cup \{\nabla \Phi_s(\mathbf{q}_\mathbf{l}) + \mathbf{l}\delta p\} \tag{4.2}$$

for $\mathbf{l} = (l_1, l_2, l_3)$ and $\mathbf{l}_k = 1$ to Np. Otherwise the p-mesh must be taken large enough to capture the maxima of the gradient of the phase functions. In either case when constructing the weights $\psi_b^{\tilde{w}}$ the $\mathbf{y}$-mesh and $\mathbf{p}$-mesh can be considered conjugate and the FFT can be used. Typically Nq is of $\mathcal{O}(\sqrt{k})$ and Np is $\mathcal{O}(1)$ [60]. Solve the ODE

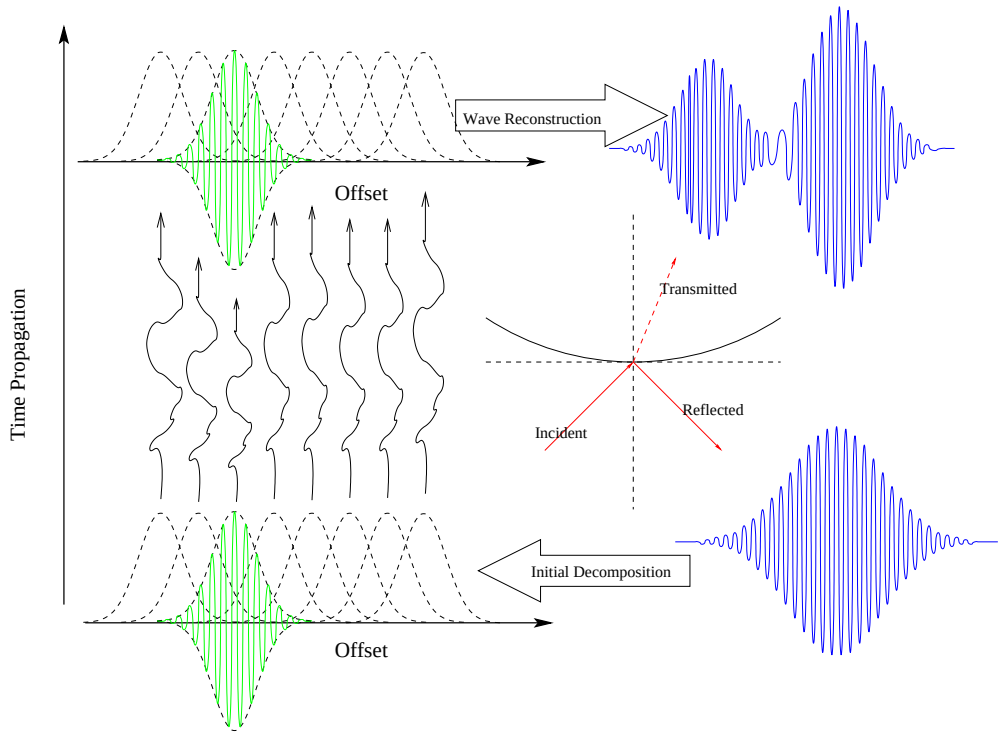


Figure 4.1: Illustration of the main three steps of the FGA algorithm: Step 1, decompose the initial wavefields into Gaussian wave packets, and calculate the corresponding weight function $\psi^k_{\mathrm{p,s}}$ defined in eq. (2.37); Step 2, Time propagation, i.e., solve numerically eqs. (2.14), (2.66), (2.34), (2.35) and (2.36) for the dynamics of Gassian center, propagation vector and prefactor amplitude. Transmission and reflection conditions are needed at the presence of interfaces as given in Section 3.2; Step 3, Wave reconstruction, i.e., to compute eq. (2.29) for the elastic wavefield at time T.

system for each Gaussian wavelet $(\mathbf{q^j}, \mathbf{p^{j,l}})$. This includes solving for $\mathbf{a}_{w,b}$ using eqs. (2.34)-(2.36), solving for $(\mathbf{Q^j}, \mathbf{P^{j,l}})$ using the flow (2.14) and solving for $Z_{w,b}$ using (2.66).
A direct sum can be used, which is akin to the trapezoidal rule, if a slice or a time-travel kernel is required. Otherwise, the fast Gaussian summation, which projects the propagated Gaussian wave packets $(\mathbf{Q}^j, \mathbf{P}^{j,l})$ to a uniform grid so that the FFT can be invoked [11].

4.2 3D Models

In this section results for are given for 3 models. First, a homogeneous media with absorbing boundary conditions. Second, the smoothed Marmousi model [9] and finally a 1-D Earth layered model IASP91 [47] For the first two examples the initial condition off of the Greens function using a forcing term (3.37); that is,

$$\rho\partial_{tt}\mathbf{u} = (\lambda+\mu)\nabla(\nabla\cdot\mathbf{u}) + \mu\Delta\mathbf{u} + \mathbf{F}(t)\delta_{\mathbf{x_0}} \tag{4.3}$$

with $\mathbf{F} = (F_1(t), F_2(t), F_3(t))$ and $\mathbf{x}_0 = (x_{1,0}, x_{2,0}, x_{3,0})$ being the spacial location of the source; see section 3.4 for formulation. For homogeneous media, the results from the FGA can be compared to SPECFEM3D [53] and the exact solution. For the forcing term we use

$$F_j(t) = \cos\left(2\pi f(t+T_0)\right)\exp\left(-(t+T_0)^2/\sigma^2\right) \tag{4.4}$$

For the ODE solver a fourth order Runge-Kutta with fixed step-size for the equations (2.14), (2.66), (2.34)-(2.36). Benchmark simulations are done on a cluster of 4 nodes each equipped with an Intel(R) Xeon(R) E5-2670 (2.60GHz) and a total of 64GB RAM.

Otherwise, all for non benchmark simulations are either done on *knot*[1] or *pod*[2] at the center for scientific computing at UC Santa Barbara.

4.2.1 Homogeneous media

For the benchmarking tests, the P- and S-wave speeds are set as $c_{\mathrm{p}} = 8$ km/s, $c_{\mathrm{s}} = 4.619$ km/s, density $\rho = 1$ kg/km^3, final propagation time $T = 13.86$ s, the computational domain is a box $[0, 128]^3$ km^3 and the source location $\boldsymbol{x}_0 = (64, 64, 64)$ km, the parameters T_0, σ are set to $T_0 = 0.1768$ s and $\sigma = 0.8660$ s. To compare the performance, the FGA is compared with the spectral element software package SPECFEM3D[3]. The accuracy and parallelizability of FGA are tested for $f = 1.4702$ Hz, and the comparison of computational time to SPECFEM3D is done for a range of f from 0.3676 Hz to 11.7617 Hz.

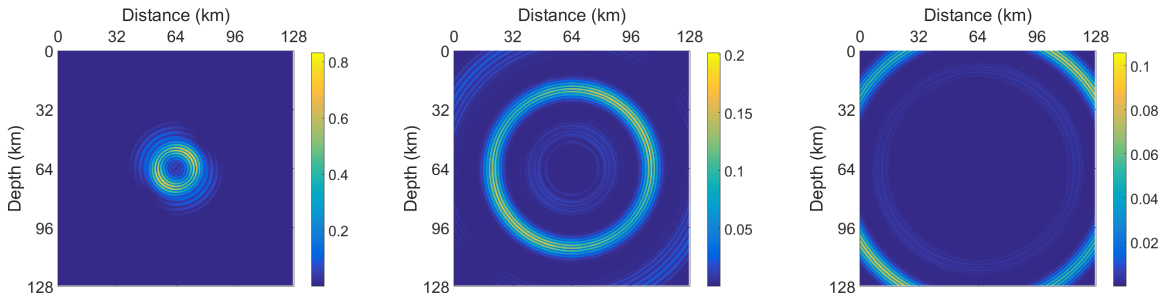
The dynamics of the elastic wave propagation is shown by the FGA method in Fig. 4.2, which is consistent with the analytical solution (3.48) as an expanding ball. The component-wise wavefields, including both P- and S-wave components, are shown in Fig. 4.3 for $t = 6.93$ s. With the comparable accuracy of FGA to SPECFEM3D as illustrated in Fig. 4.4 for $t = 6.93$ s and the source frequency $f = 1.4702$ Hz. The relative error for Fig. 4.4(d), FGA versus the analytical solution is computed to be 3.84%; while for Fig. 4.4(e), the relative error is computed to be 3.57%. With a comparable accuracy, the FGA shows a much faster computational speed and better parallelizablity than SPECFEM3D for high-frequency (≥ 1.4702 Hz) elastic wave propagation, with details described in Figs. 4.5 and 4.6. Particularly, one can see in Fig. 4.5 that the computational time of SPECFEM3D has nearly cubic growth in the frequency of elastic waves, while

[1] http://csc.cnsi.ucsb.edu/clusters/knot
[2] http://csc.cnsi.ucsb.edu/clusters/pod
[3] https://geodynamics.org/cig/software/specfem3d/

FGA has about 1.5-times growth in the frequency. Fig. 4.6 shows that the speed-up ratio for FGA is almost 2 when one doubles the number of processors, which indicates that the FGA algorithm is embarrassingly parallel. Remark that, in the simulation for the source frequency $f = 1.4702$ Hz where FGA and SPECFEM3D produce comparable accuracy, we use 1504436 Gaussians for computing P-wave of both "$\pm$" propagation directions, and 2120482 Gaussians for computing S-wave of both "$\pm$" propagation directions, which needs to roughly compute a total number of $8 \times (1504436 + 2120482) \approx 30$ millions of variables in the simulation. Note that, the equations for both P- and S-waves have the same number of variables, and the prefactor 8 comes by counting the number of variables needed in eqs. (2.14), (2.34)-(2.36) where $\boldsymbol{Q}$ and $\boldsymbol{P}$ are 3-D real vectors and the prefactor amplitude a is a complex number; while in SPECFEM3D, we use 128 elements in each direction with 5^3 nodes in each element. One needs to roughly to compute a total number of $3 \times 128^3 \times 5^3 \approx 800$ millions of variables in the simulation, where the prefactor 3 is due to $\boldsymbol{u}$ is a 3-D real vector in eq. (2.2). In addition, the stability conditions on time step for solving the ODE systems (2.14) and (2.34)-(2.36) by RK4 is better than solving the elastic wave equation (2.2) by SPECFEM3D, since the CFL condition is restricted by small wavelength in solving eq. (2.2).

4.2.2 Smoothed Marmousi model

The Marmousi model based on a profile of the North Quenguela through the Cuanza basin in Angola [9]. We consider a smoothed version of this [104]. We use the data for the P-wave velocities and for the S-wave we divide by the factor of $\sqrt{3}$. The computational domain is $[4, 8.032]$ km by $[0, 2.016]$ km by $[0, 2.904]$ km . As the model is given for a slice, the velocities as constant across the y-direction. We take $T_0 = 0.125$ s, $\sigma = 0.0442$ s, $f = 40.7436$ Hz.



(a) Initial wavefield (b) $t = 6.93$ s (c) $t = 13.86$ s

Figure 4.2: Modulus of the elastic wavefield computed by the FGA method in homogeneous media, with the analytical solution given by eq. (3.48). The subfigures from left to right show the slices of $\|\boldsymbol{u}\|_2$ at $y = 64$ km for $t = 0, 6.93, 13.86$ s. The frequency of the source time function is $f = 1.4702$ Hz.

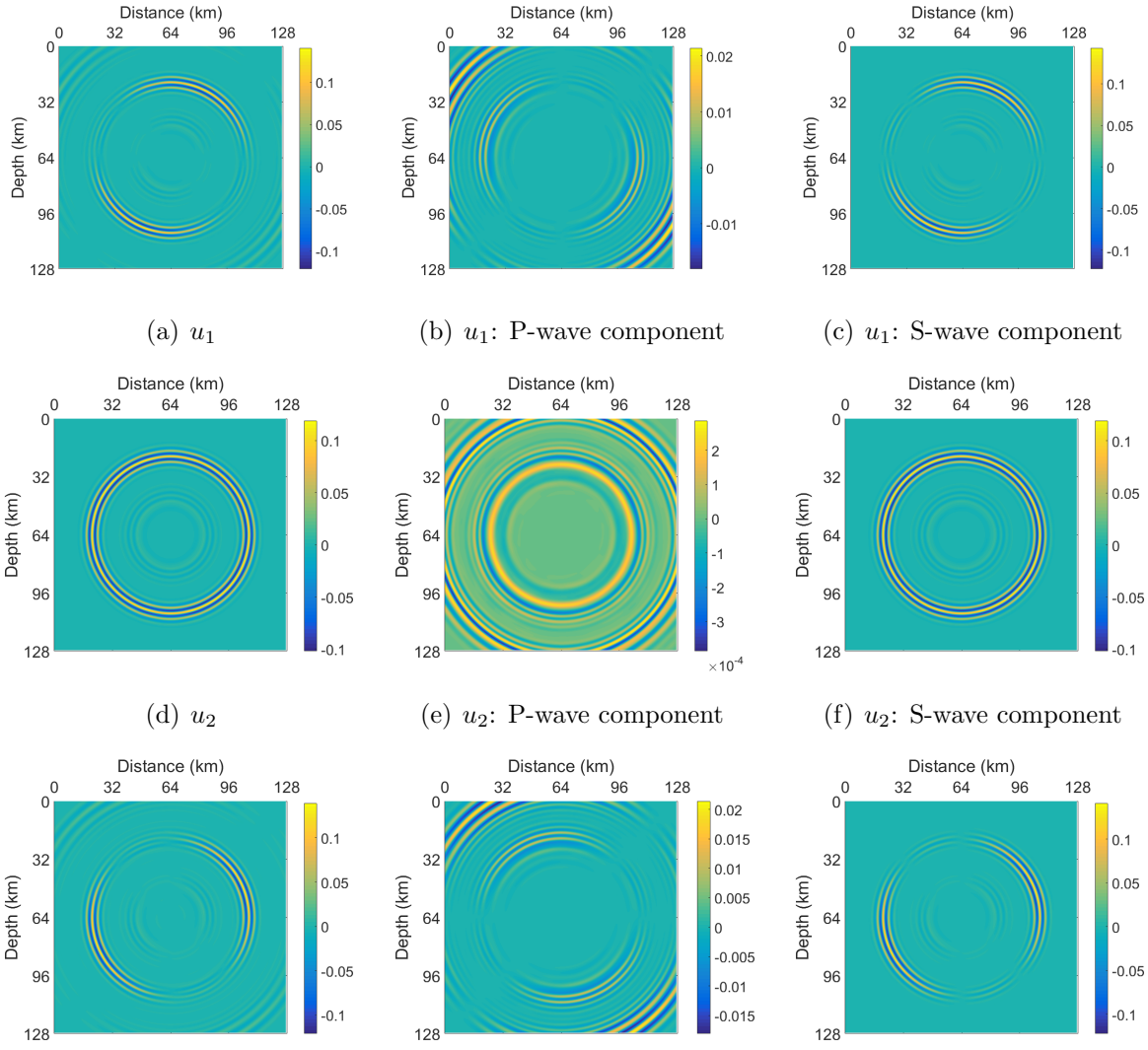
4.2.3 Waveguide example in a 1-D layered Earth model

We verify the interface conditions (3.28) and (3.24) by simulating a waveguide example in a 1-D layered Earth model, with the layered P-wave velocity given in Fig. 4.10(a) following the data in the IASP91 model [51, 47]. We are particularly interested in choosing the 410-km discontinuity for the numerical proof of the conditions (3.28) and (3.24), which presents to a $5-6\%$ increase on P-wave velocity and calibrates the mantle transition zone. We consider a radially symmetric surface source as shown in Fig. 4.10(b), so that the elastic wave equation (2.2) has a solution of P-waves in the form of $\mathbf{u}(t, \mathbf{x}) = \nabla\psi(t, |\mathbf{x}|)$ where $\psi(t, r)$ is the radially symmetric solution to the scalar wave equation, i.e.,

$$\partial_t^2\psi - c_{\mathrm{p}}^2(\boldsymbol{x})\Big(\partial_r^2\psi + \frac{2}{r}\partial_r\psi\Big) = 0. \tag{4.5}$$

We choose a setup that allows radially symmetric solution so that the compared solution can be solving by the following 1-D wave equation along the radius direction with fine enough grid points,

$$\partial_t^2 u - c_{\mathrm{p}}^2(\boldsymbol{x})\Big(\partial_r^2 u + \frac{2}{r}\partial_r u\Big) = 0.$$

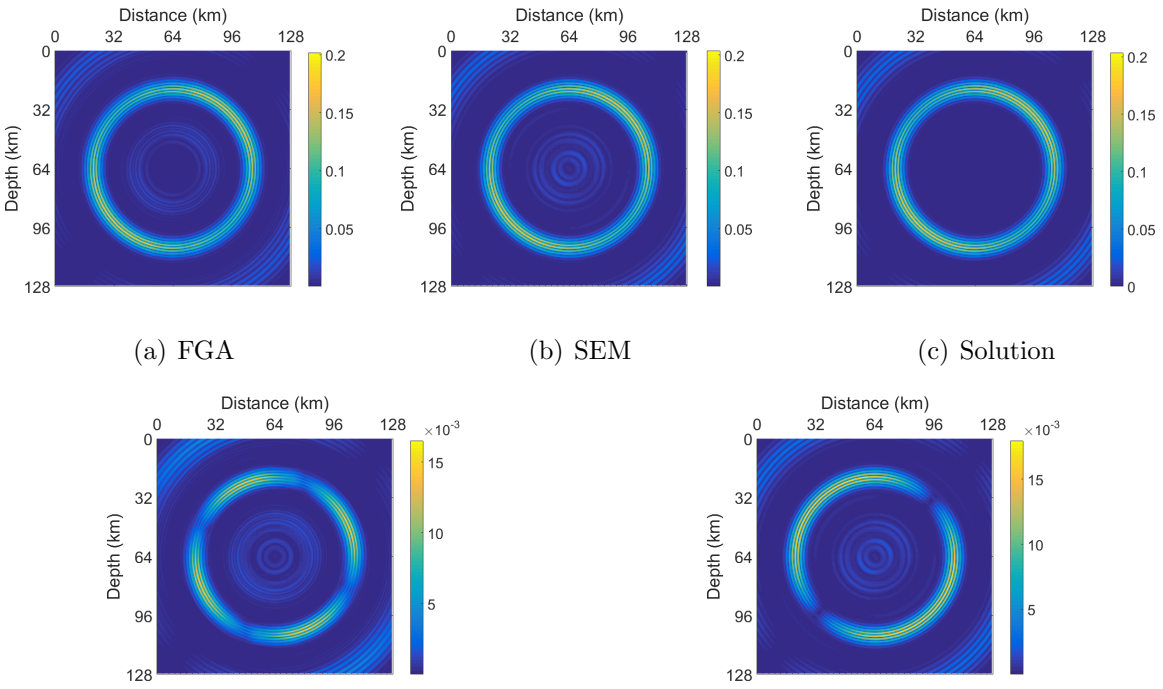


(a) u_1 (b) u_1: P-wave component (c) u_1: S-wave component

(d) u_2 (e) u_2: P-wave component (f) u_2: S-wave component

(g) u_3 (h) u_3: P-wave component (i) u_3: S-wave component

Figure 4.3: Modulus of the component-wise elastic wavefield computed by the FGA method in homogeneous media, with the analytical solution given by eq. (3.48). The subfigures show the components of $\boldsymbol{u} = (u_1, u_2, u_3)^{\mathrm{T}}$ for $t = 6.93$ s and the source frequency $f = 1.4702$ Hz. The second and third columns show the computed P- and S-wavefields by the FGA method, respectively.



(a) FGA (b) SEM (c) Solution

(d) FGA vs Solution (e) SEM vs Solution

Figure 4.4: Comparison of accuracy for FGA and SPECFEM3D to the analytical solution (3.48) at $t = 6.93$ s with the source frequency $f = 1.4702$ Hz. The top subfigures are the modulus of wavefields given by FGA, SPECFEM3D and analytical solution. The bottom subfigures are the modulus differences between the FGA and analytical solutions (bottom left) and between the SPECFEM and analytical solutions (bottom right). This shows that FGA and SPECFEM3D produce comparable accuracy in this case.

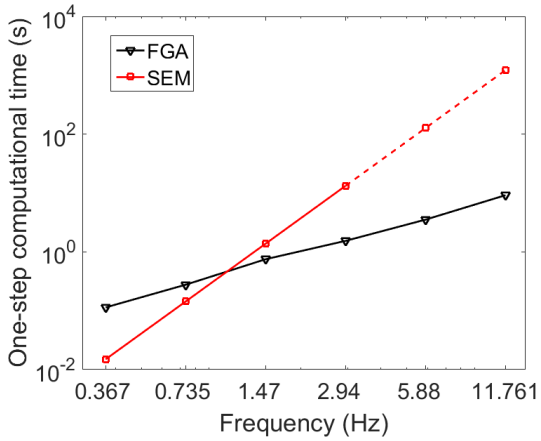


Figure 4.5: Dependence of one-step computational time on frequency for both FGA and SPECFEM3D in homogeneous media. The horizontal axis is the frequency f (Hz) and the vertical axis is the one-step computational time (s) of the solvers. The triangle line stands for the FGA simulations and the square line stands for the SPECFEM3D simulations. Due to the limitation of memory, SPEFEM3D can not run for $f \geq 5.88084$ Hz, and the dashed square line is obtained by extrapolation.

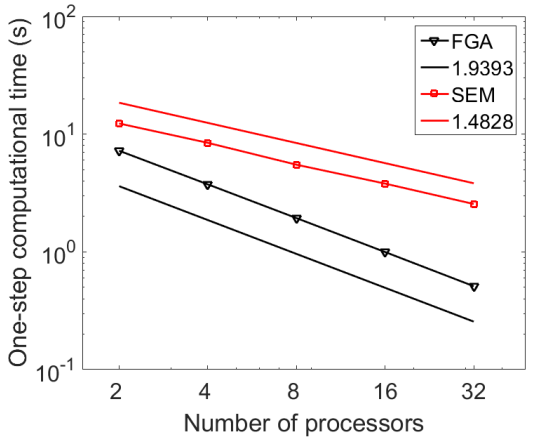
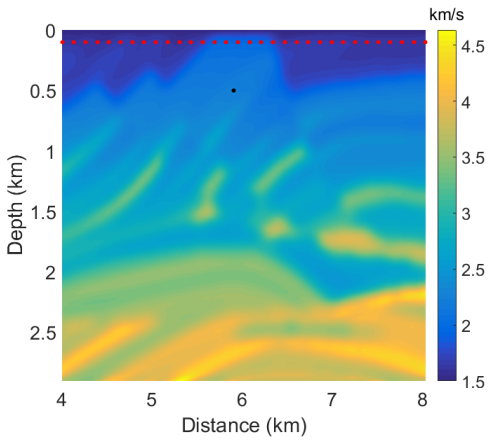
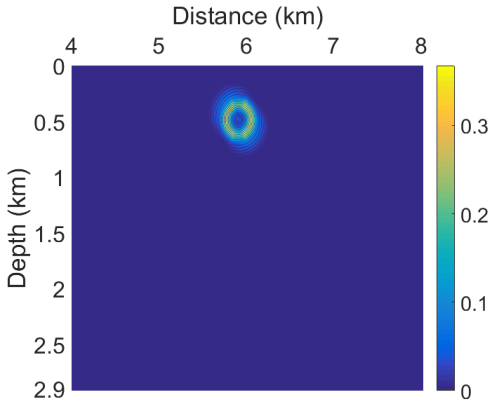


Figure 4.6: Dependence of time on the number of processors for the source frequency $f = 1.4702$ Hz in homogeneous media. The ideal speed-up ratio is 2, while one can see that the speed-up ratio for FGA is approximately 1.9393, which is slightly smaller than 2 indicating an almost perfectly parallel efficiency. On the other hand, as a comparison, SPECFEM3D is used with 128 elements in each spatial direction to achieve a comparable accuracy to FGA. The speed-up ratio for the SPECFEM3D solver is around 1.4828, which is smaller that those of FGA. This is because SPECFEM3D solves eq. (2.2) on a parallel computer with N processors by partitioning the whole domain into N slabs with each processor solving the equation in each slab. Therefore, for each time step, each processor needs to communicate with its neighbors to get necessary boundary information, which decreases the speed-up ratio.



(a) P-wave velocity profile.



(b) Modulus of initial wavefield

Figure 4.7: The Marmousi Model (a) P-wave velocity profile for an xz-plane. Stations are in red spanning horizontally at 130 m intervals starting at $(4, 1.008, 0.1)$km. Source is located in black at $(5.9, 1.008, 0.5)$ km. The velocities are in km/s (b) Slice of the Marmousi Model at $y = 1.008$ km with $T_0 = 0.125s$, $\sigma = 0.0442$ s, $f = 40.7436$ Hz

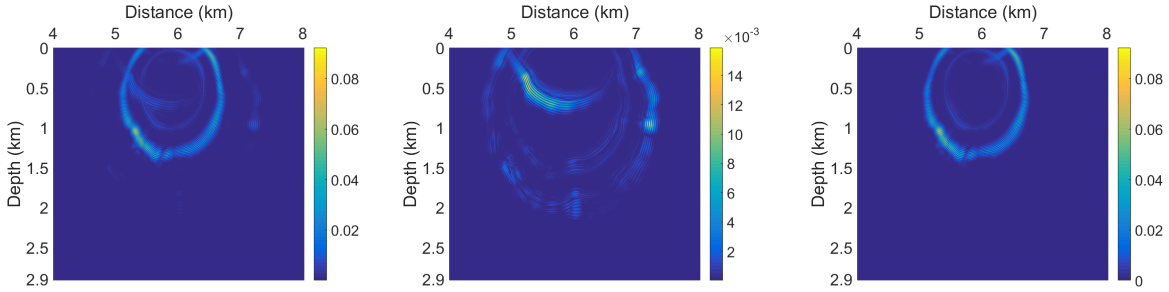
We choose the initial condition for the elastic wave equation (2.2) as

$$\mathbf{u}(0, \mathbf{x}) = \nabla\phi_0(r - r_0), \quad \text{and} \quad \partial_t \mathbf{u}(0, \mathbf{x}) = \mathbf{0},$$

where $r = |\mathbf{x}|$, $r_0 = 600$ km, and

$$\phi_0(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right)\cos\left(\frac{2\pi r}{\ell}\right),$$

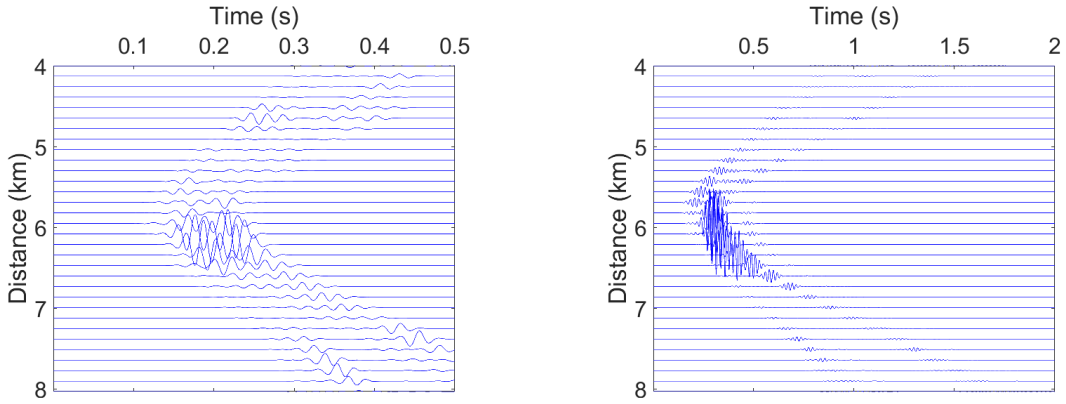
with $\sigma = 3.3146$ km and $\ell = 7.3631$ km. The dominant frequency is around 1.36 Hz. We solve eq. (4.5) using 1-D finite difference method with fine enough grid points as the reference solution, to which we compare the full elastic wave solution of (2.2) computed by the FGA algorithm. Fig. 4.11 shows the seismic signals of P-waves received at stations of depth 480 km, 420 km and 360 km, respectively, where one can see a good agreement of FGA simulation with the reference solution for the P-waves.



(a) Modulus of wavefield at $t = 0.5$ s (b) Modulus of P-wavefield at $t = 0.5$ s (c) Modulus of S-wavefield at $t = 0.5$ s

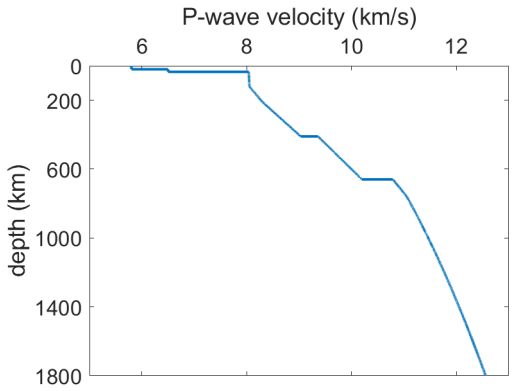
(d) Modulus of wavefield at $t = 1$ s (e) Modulus of wavefield at $t = 1.5$ s

Figure 4.8: Modulus of P and S wavefields at various time slices. Top row: $t = 0.5$ s (a) Modulus of wavefield. (b) Modulus of P-wavefield. (c) Modulus of S-wavefield: Bottom row: (d) Modulus of wavefield at $t = 1$ s. (e) Modulus of wavefield at $t = 1.5$ s

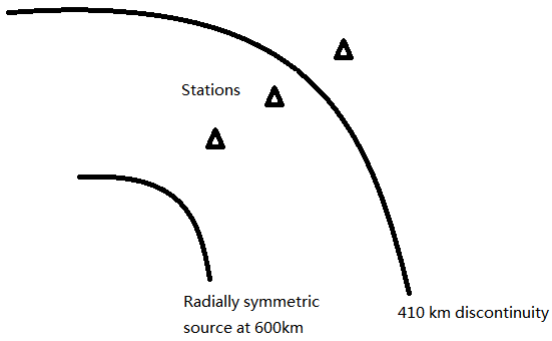


(a) Seismogram at $t = 0.5s$ (b) Seismogram at $t = 2s$

Figure 4.9: Seismograms for $T_0 = 0.125s$, $\sigma = 0.0442s$, $f = 40.7436$Hz, receivers located at 130m intervals.

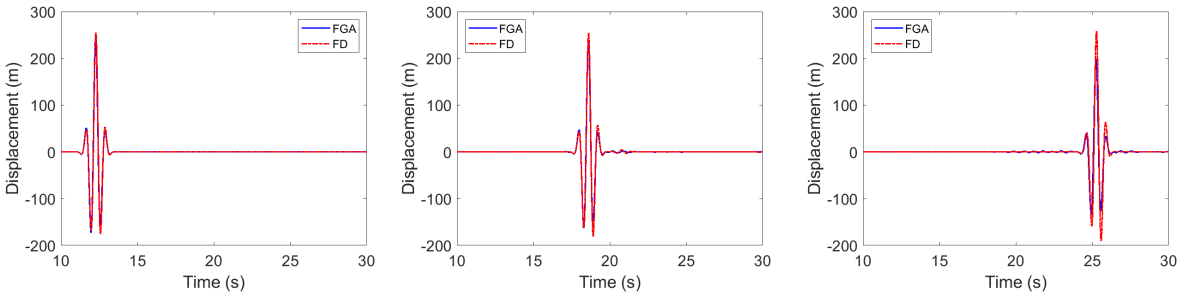


(a)



(b)

Figure 4.10: (a): The layered P-wave velocity follows the data in the IASP91 model [51, 47]. (b): Cartoon plot of the 410-km discontinuity, and distributions of source and stations.



(a) P-wave at depth 480 km (b) P-wave at depth 420 km (c) P-wave at depth 360 km

Figure 4.11: The seismic signals of P-waves received at stations of depth 480 km, 420 km and 360 km, respectively, simulated by the finite difference (FD) and FGA methods, with the source locating at the depth of 600 km. FGA shows a good agreement with the reference signals computed by the FD method for the P-waves.

4.3 Seismic Tomography

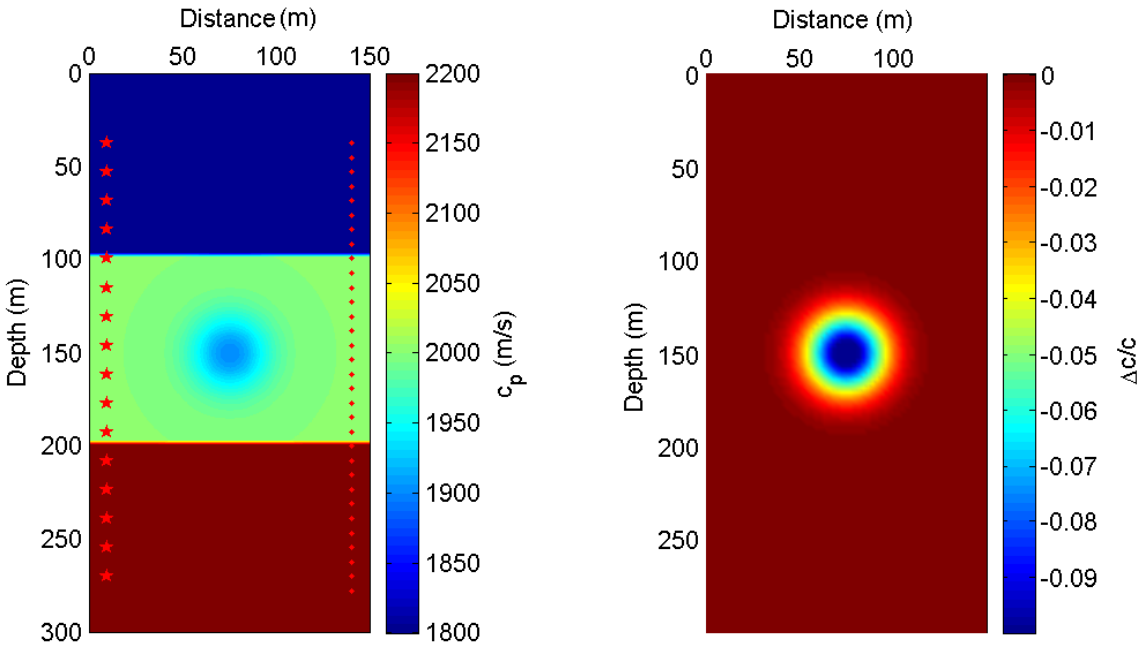
In this section, the FGA for high frequency elastic waves is applied to 3-D wave-equation-based traveltime tomography [91, 56] and FWI [71, 96]. Also, optimal transport theory-based seismic tomography proposed in [102] is investigated. Other strategies for optimal transport theory-based seismic tomography; as in [62], require further studies.

4.3.1 Three-layered crosswell model

In particular, we consider the following crosswell seismic tomography, where the three-layered velocity model is set up as follows, with a low-velocity region located at the second layer and homogeneity in the y-direction,

$$c_{\mathrm{p}}(x,y,z)=\sqrt{3}c_{\mathrm{s}}(x,y,z)=\begin{cases} C_1, & \text{if } z_0<z<z_1,\\ C_2\left(1-\alpha\exp^{-\beta\left((x-x_c)^2+(z-z_c)^2\right)}\right), & \text{if } z_1<z<z_2,\\ C_3, & \text{if } z>z_2, \end{cases} \tag{4.6}$$

where the layered velocities are $C_1 = 1800$ m/s, $C_2 = 2000$ m/s, $C_3 = 2200$ m/s, and the interfaces locate at $z_0 = 0$ m, $z_1 = 100$ m, $z_2 = 200$ m. The center of the low-velocity region is set at $x_c = 75$ m, $z_c = 150$ m. We choose $\alpha = 10\%$ and $\beta = 1/450$ m^{-2} to indicate the largest magnitude of the low-velocity perturbation from the background velocity and the area of low velocity region; see Fig. 4.12 for an illustration of the P-wave velocity. In Fig. 4.12, we also show the positions of 16 seismic resources (as stars) with an equal spacing of 16 m in one well and 32 seismic receivers with an equal spacing of 8 m in the other well.



(a) Three-layered crosswell model (b) Relative velocity $(c_{\mathrm{p}} - c_0)/c_0$

Figure 4.12: (a): The three-layered crosswell P-wave velocity model with homogeneity in horizontal y-direction. The background velocities from top to bottom are $C_1 = 1800$ m/s, $C_2 = 2000$ m/s, $C_3 = 2200$ m/s, and the interfaces locate at $z_0 = 0$ m, $z_1 = 100$ m, $z_2 = 200$ m. The low-velocity region has a Gaussian shape centered at $x_c = 75$ m, $z_c = 150$ m, with standard deviation equal to 1/30 m. 16 stars indicate the locations of seismic sources, and 32 dots indicate the locations of seismic receivers. (b): we use the relative velocity $\Delta c/c = (c_{\mathrm{p}} - c_0)/c_0$ to indicate the largest magnitude of the low-velocity perturbation from the background velocity and the area of low velocity region.

We start with the following initial velocity model

$$c_0(x,y,z) = \begin{cases} C_1, & \text{if } z_0 < z < z_1, \\ C_2, & \text{if } z_1 < z < z_2, \\ C_3, & \text{if } z > z_2, \end{cases} \tag{4.7}$$

then use FGA to compute the forward and adjoint wavefields and construct 3-D kernels of different phases by the methods of wave-equation-based traveltime tomography [91, 56] and FWI [71, 96]. Since FWI requires a more sophisticated initial velocity model for the convergence than wave-equation-based traveltime tomography, a hierarchical strategy from [12] is used. First wave-equation-based traveltime tomography is used to create a macro-scale model and then FWI to generate a high-resolution micro-scale model. For the signals received at the top station (Fig. 4.13), we show the corresponding kernels of different phases in Fig. 4.14. The inversion results are shown in Fig. 4.15. The damping and smoothing parameters are chosen empirically by forcing the variation in each iteration less than 3%. Note that artifacts are visible in the final images, which are unavoidable and mainly caused by the uneven data coverages.

4.3.2 Optimal Transport Theory-based Seismic Tomography

A numerical instability when one uses the adjoint-state method with 2-Wasserstein metric for FWI with high-frequency seismic data, which leads to the non-convergence of seismic tomography computed by the FGA algorithm. We illustrate this instability by a simple example described as follows. Consider f and g are two high-frequency seismic signals received at a station within the time period $[0, T]$, as illustrated in a cartoon

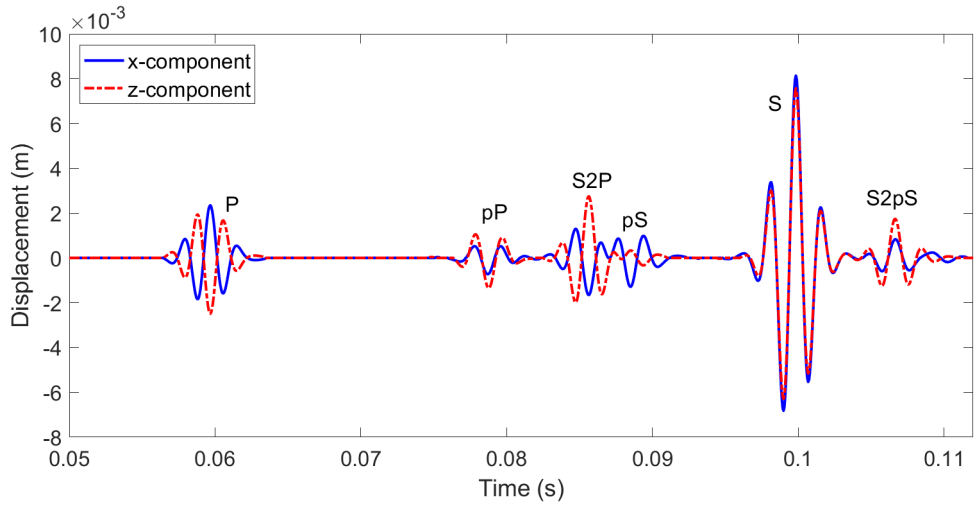


Figure 4.13: The signals of direct P, pP, S2P, pS, direct S, and S2pS received at the top station with the source as the top ninth source in Fig. 4.12. Here 2 means the second interface locates at $z_2 = 200$ m, e.g. S2P means S-wave gets reflected at the interface at $z_2 = 200$ m and arrives at the station as P-wave.

figure Fig. 4.16,

$$f(t) = \begin{cases} \sin(\omega t) & \text{if} \quad 0 \le t \le \dfrac{2\pi}{\omega}, \\ 0 & \text{otherwise}, \end{cases} \qquad g(t) = \begin{cases} \sin\big(\omega(t-d)\big) & \text{if} \quad d \le t \le \dfrac{2\pi}{\omega} + d, \\ 0 & \text{otherwise}, \end{cases}$$

where $\omega \gg 1$ indicates the received signals are of high-frequency.

Following the strategy proposed in [102], the normalized two distributions corresponding to the signals f and g are

$$P(f) = \begin{cases} \dfrac{1+\sin(\omega t)}{T} & \text{if} \quad 0 \le t \le \dfrac{2\pi}{\omega}, \\ \dfrac{1}{T} & \text{otherwise}, \end{cases} \qquad P(g) = \begin{cases} \dfrac{1+\sin\big(\omega(t-d)\big)}{T} & \text{if} \quad d \le t \le \dfrac{2\pi}{\omega} + d, \\ \dfrac{1}{T} & \text{otherwise}, \end{cases}$$

where we have chosen to add a constant 1 to make signals positive, and normalize by a factor of $1/T$.

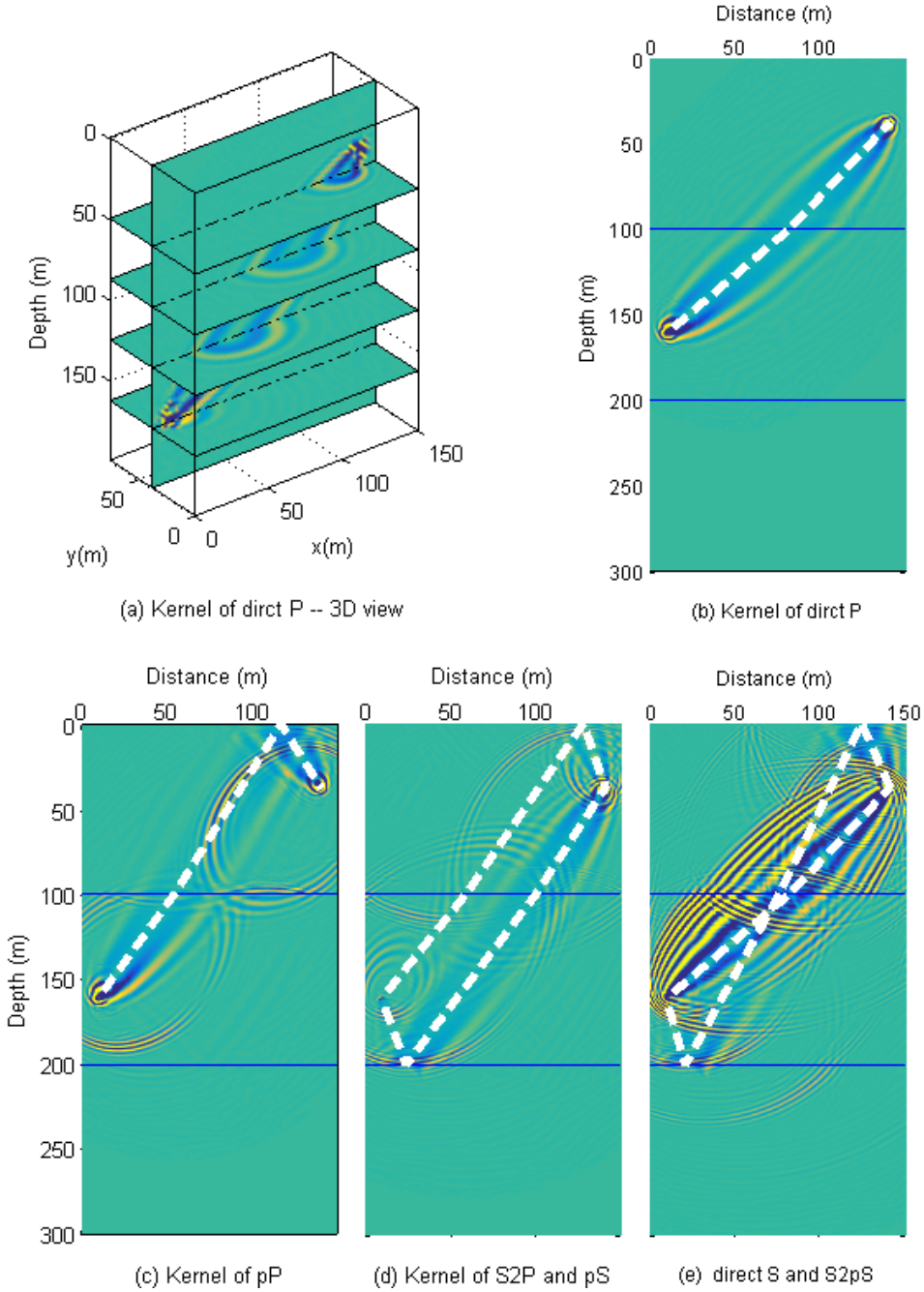


Figure 4.14: Seismic tomography kernels computed by FGA for the initial velocity model (4.7), with the thick dashed lines as the actual ray paths of direct P, pP, S2P and pS, direct S, and S2pS signals in Fig. 4.13, respectively. (a): Kernel computed from the direct P signal – 3D slices view; (b): Kernel computed from the direct P signal; (c): Kernel computed from the pP signal; (d): Kernel computed from the S2P and pS signals; (e): Kernel computed from the direct S and S2pS signals.

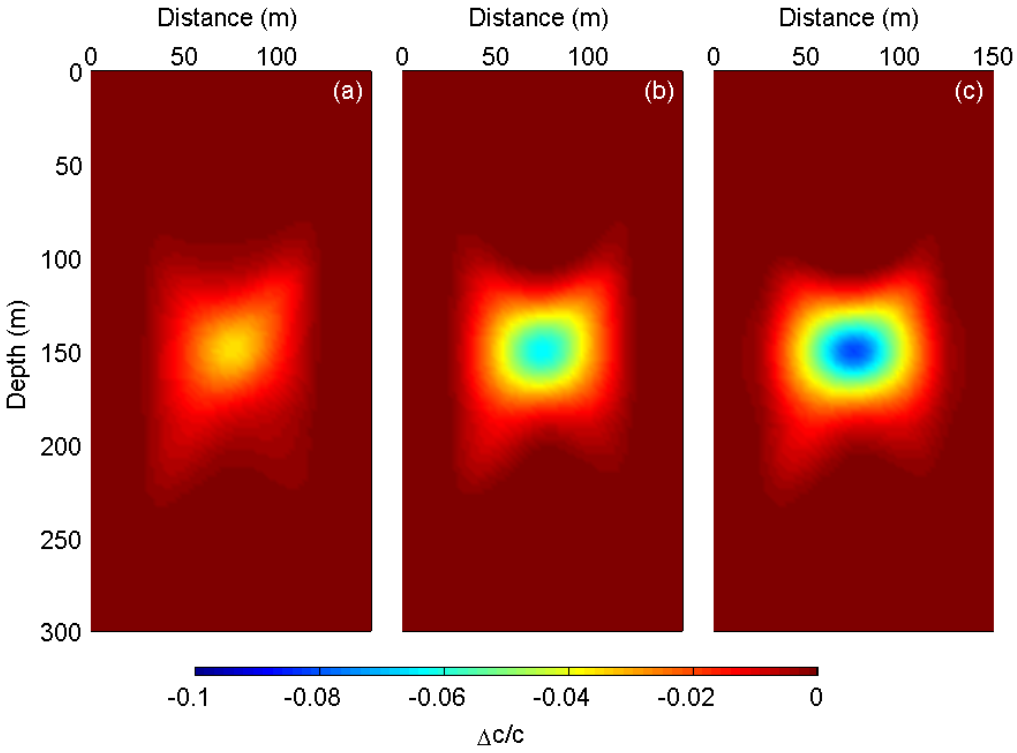


Figure 4.15: For $\alpha = 10\%$ and $\beta = 1/450$ m^{-2} in eq. (4.6), the subfigures (a) and (b) are the first two iterations using travel-time tomography, and the subfigure (c) is the third iteration using FWI which removes the artifacts in travel-time tomography and is close to the true velocity profile given in Fig. 4.12(b).

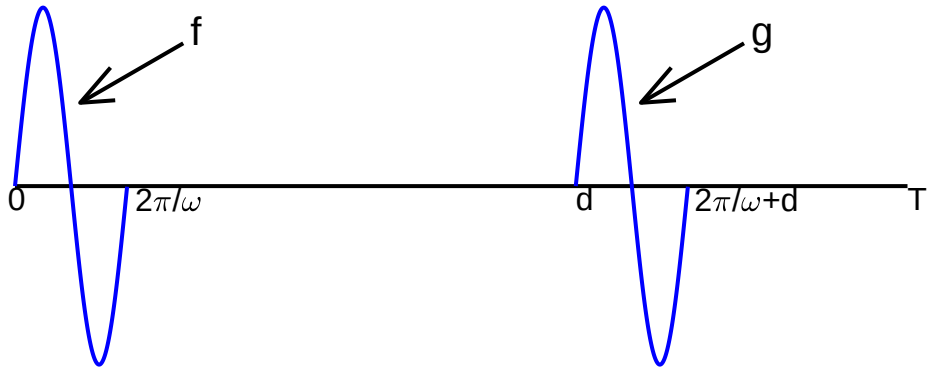


Figure 4.16: Two successive high-frequency ($\omega \gg 1$) seismic signals received at one station within the time period $[0, T]$. The distance between f and g is d, which implies the corresponding Wasserstein-2 metric is equal to d.

Then the accumulative distribution functions are given by integrating $P(f)$ and $P(g)$,

$$F(t) = \begin{cases} \dfrac{t}{T} + \dfrac{1-\cos(\omega t)}{\omega T} & \text{if} \quad 0 \le t \le \dfrac{2\pi}{\omega}, \\ \dfrac{t}{T} & \text{otherwise}, \end{cases}$$

$$G(t) = \begin{cases} \dfrac{t}{T} + \dfrac{1-\cos\big(\omega(t-d)\big)}{\omega T} & \text{if} \quad d \le t \le \dfrac{2\pi}{\omega} + d, \\ \dfrac{t}{T} & \text{otherwise}, \end{cases}$$

which implies

$$F(t) = \frac{t}{T} + \mathcal{O}(\omega^{-1}), \quad G(t) = \frac{t}{T} + \mathcal{O}(\omega^{-1}), \quad \text{when} \quad \omega \gg 1. \tag{4.8}$$

Therefore, by eq. (9) in [102], one has

$$W_2^2(f,g) = \int_0^1 |t - G^{-1}\big(F(t)\big)| f(t)\,\mathrm{d}t = \mathcal{O}(\omega^{-1}) \ll 1,$$

where we have used the fact that $G^{-1}\big(F(t)\big) = t + \mathcal{O}(\omega^{-1})$ by eq. (4.8). However, on the other hand, by the definition of the wasserstein-2 metric, one has $W_2(f,g) = d$ as the translate distance of moving the signal f to coincide with the signal g. This inconsistency shows that the two successive high-frequency signals of distance d will be treated as two closely arriving signals in Wasserstein-2 metric, which leads to the nonconvergence of seismic tomography computed by the FGA algorithm. This is due to an existence of numerical instability, and may be resolved if, for example, one uses extremely small time steps in SPECFEM3D or increase the asymptotic accuracy of FGA to be of order $\mathcal{O}(\omega^{-2})$ or higher, which would impose a bottleneck of computational time when one simulates the high-frequency elastic wave propagation in 3-D.

Chapter 5

Deep Leaning with the FGA

In this chapter a deep learning algorithm for seismic interface and pocket detection with neural networks trained by synthetic high-frequency displacement data efficiently generated by the FGA is proposed and reported on. In seismic imaging, high-frequency data is advantageous since it can provide high resolution of substructures. However; as discussed in 3.1, generation of sufficient synthetic high-frequency data sets for training neural networks is computationally challenging. This bottleneck is overcome by a highly scalable computational platform built upon the FGA, which is verified in section 4.2. Neural networks usually excel in is to classify the objects based on what they have been trained, and they may do a terrible job to recognize an object which does not fall in the categories of training set but obviously belong to one of the categories by human intuition. Typically, neural networks are designed and used for classification problems. Once trained the network might fail to categorize accurately, compared to human intuition, an object outside the convex hull of parameters from its training dataset. Although it cannot be certain what a neural network learns from its training, we aim incorporate some "intuition", the concept of traveltime, into the training of neural networks for seismic interface detection and pocket detection. These data sets contains accurate traveltime

information from the ray path; but not exact amplitude information, i.e asymptotic errors not shrinking to zero even at extremely fine numerical resolution. Using this data we build convolutional neural network models using an open source API, GeoSeg, developed using Keras and Tensorflow. On a simple model, networks, despite only being trained on FGA data, can detect an interface with a high success rate from displacement data generated by SPECFEM3D. This is illustrated by considering a simple two-dimensional layered velocity model where, apparent to human beings, traveltime matters more than other factors in detecting the interface. Benchmark tests are done for P-waves (acoustic) and P- and S-waves (elastic) generated using the FGA and a spectral element method. Further, results with a high accuracy are shown for more complicated geometries including a three-layered model, and a 2D-pocket model where the neural networks trained by both clean and noisy data.

5.1 Neural Networks and seismology

5.1.1 Introduction

Neural networks excel at recognizing shapes, patterns, and sorting relevant from irrelevant data; this makes them good for image recognition and classification. In particular, convolutional neural networks allowed for rapid advances in image classification and object detection [54], and in fact networks have been created for specific tasks, such as, fault detection [4], earthquake detection, *ConvNetQuake* [68], *DeepDetect* [99] and seismic phase arrival times, *PhaseNet* [107]. A CMP stack or seismogram can be thought of as an image from a seismic event in which information about subsurface structures is stored. In terms of reflective seismology a seismogram records information and the refractive index of an interface is encoded in that seismogram. Seismic migration is a

process by which seismic events are geometrically re-located to their incidence location, rather than the recorded location, in the subsurface. The goal of a network should be to decode this information.

One obstacle in building a neural network to detect seismic structures is having an ample data set for training. There is constant waveform data being collected by seismic stations across the globe, and generating data by resampling of this seismic data to train a network can be done, but is limited by the Nyquist frequency. Seismic data can not be resampled with a Nyquist frequency lower than the highest usable frequency in the data, thus high frequency data is usually preferred as it tends to lead to improved resolution of the substructures. Other difficulties of gathering an ample data lie within the differences in geological locations, natural phenomenon (e.g. earthquakes) and unnatural phenomenon (e.g. fracking). Using these data sets to train a general neural network is a daunting task, and thus it is natural to use synthetic data for the training of neural networks. The dominant frequency of a typical earthquake is around 5 Hz [65] leading to demanding, and at times, unaffordable computational cost. This makes generation of sufficient synthetic high-frequency data sets for training neural networks computationally challenging with well-known methods discussed in 3.1. As the FGA algorithm is embarrassingly parallel, this difficulty is overcome.

Using synthetic data, Araya-Polo et al. perform inverse tomography via fully connected neural networks with great success in [5] . Their networks use low dimensional features extracted from seismic data as input. Using deeper convolutional neural networks trained on seismogram data may allow the network to pick up on previously unknown signals. The increase in input dimensionality necessitates more sophisticated deep learning techniques than those presented in [5].

The data points used for our experiments are generated from the forward simulation of the elastic wave equation using the FGA. We record the displacement data from the wavefield at receiver locations. Given an initial condition, the initial wave packet decomposition can be saved for a variety of tests. This means the same data can be loaded as the parameters vary from data point to data point. If the initial condition is independent of the wave velocities, the same initial wave packet decomposition can be used to generate seismograms with varying velocities, and varying interface depth. Hence for the forward simulation, loading the initial wave packet decomposition, running an ODE solver, and recording the seismograms are the only tasks required. As the ODE system for the FGA is uncoupled for each wave packet, the speed of a single simulation greatly benefits from a parallel implementation.

5.1.2 Network Design

As reported in section 4.3, the goal of FWI is to extract wave speed data from seismic data. In its purest form, this is a regression type problem and was addressed with fully connected networks in [5]. Our work approaches the problem from a segmentation perspective. We address a simplified version of FWI and attempt to detect subsurface structures by classifying them as regions of low or high wavespeed, thus transforming the regression problem into a segmentation problem. These sorts of segmentation problems have been addressed with great success by CNNs [81]. Semantic segmentation of images is the process of labeling each pixel in an image with a class label for which it belongs. In semantic segmentation problems the correct pixel label map is referred to as the ground truth. In our work the "image" is the n-dimensional slice in the depth direction which we normalize and partition into N bins which act as our "pixels". Each bin is then labeled depending on whether it came from a region of high or low velocity. These velocity regions

are our classes. Our work diverges substantially from traditional semantic segmentation of images, as our input is time series data which must be transformed by the network. This is opposed to the traditional case where the input itself is labeled. The goal of our network is to infer the presence of high and low wavespeed regions and the interfaces between them from seismogram data. The input to the network is $X \in \mathbb{R}^{M \times d \times r}$, where M is the number of timesteps, d is the spatial dimension of media, and r is the the number of receivers. The output of the network is

$$\mathcal{N}(X) = (p^k_{i_1 \dots i_n}) \in \mathbb{R}^{M_1 \times \cdots \times M_n \times N}, \qquad \begin{matrix} i_j \in \{1, ..., M_j\} \\ k \in \{1, \dots, N\} \end{matrix}, \tag{5.1}$$

where $p^k_{i_1 \cdots i_n}$ is the probability that bin $i_1 \cdots i_n$ belongs to the k^{th} class. In this paper $d = 3$, $n = 1, 2$, and $N = 1, 2, 3$. The accuracy of a given inference is found by taking the argmax along the last axis of the output tensor and comparing against the groundtruth. Taking a max along the last axis recovers the probability, interpreted as a confidence, of the prediction. We call this value the heatmap. In [5], Araya-Polo et al. perform inverse tomography via Deep Learning and achieve impressive results. Our model is fundamentally different than GeoDNN in that: GeoDNN is a fully connected network whereas GeoSeg's is fully convolutional, and GeoDNN uses semblance panels from CMP data as features for the network and GeoSeg uses the raw seismograph data. Moreover, Araya-Polo et al. address the FWI problem and provide the wave speeds in a two dimensional region and we tackle high and low velocity detection, shifting the problem from regression to segmentation.

The networks were built using an open source API, GeoSeg[1], developed using Keras and Tensorflow. GeoSeg supports UNet, fully convolutional segmentation network, or

[1] https://github.com/KyleMylonakis/GeoSeg

feed forward CNNs as a base meta-architecture, using any of residual, dense, or convolutional blocks, with or without batch normalization [78, 81, 37, 44, 46]. GeoSeg also allows for easy hyper-parameter selection for network and block architectures, and for training optimizers and parameters. The optimizers used were NADAM with default parameters [21], sometimes followed by minibatch stochastic gradient descent (SGD), or SGD alone. The network structures are described by their meta-architecture and their blocks. The meta-architecture describes the global topology of the network and how the blocks interact with each-other. Each block either begins or ends with a decoding or encoding transition layer respectively. Encoding transition layers downsample their inputs with a strided convolution. Decoding transition layers upsample their inputs with a strided deconvolution. Transition layers will not have dropout.

Meta-Architectures. While GeoSeg supports many kinds of feed-forward CNN's and Encoder-Decoder Networks with different choices of blocks, UNet architectures with dense blocks performed the best and will be the only type of network reported.

GeoDUDe-L refers to a UNet architecture from [78]. These architectures have proven highly efficient at image segmentation for road detection [105] and in biomedical applications [78]. These networks feed their input into a transfer branch, then an encoder branch of length L, bridge block, and then a decoder branch of length L. The last layer is a convolutional layer followed by a softmax which outputs predictions as described above. The defining feature of these networks are the "rungs" connecting the encoder and decoder branches (see Figure 5.1). In this way, the network can incorporate both low and high resolution data [78, 105]. For the one dimensional problems the transfer branch is not necessary and can be omitted.

Convolutional Layers. The layer is broken first into a bottleneck convolution followed by the main convolution. The bottleneck is a convolution which uses a 1x1 kernel to expand the number of feature channels before performing the full convolution.

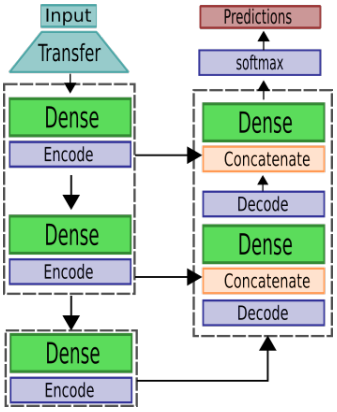
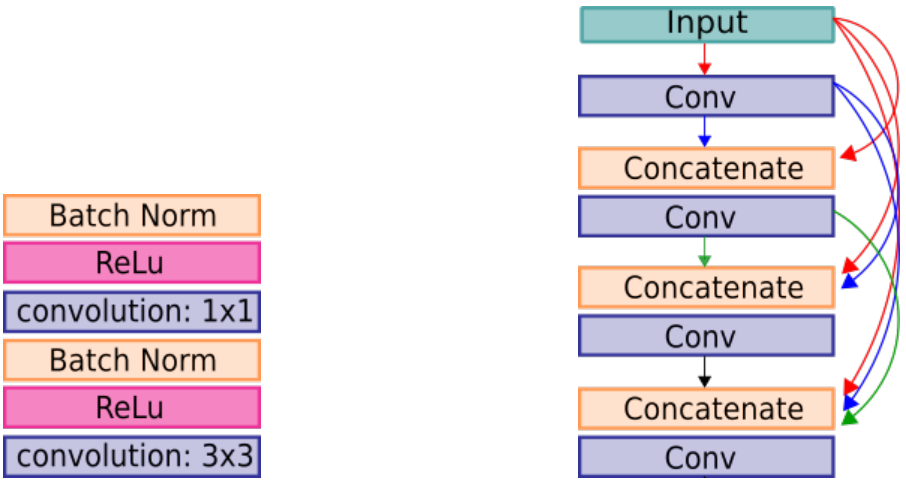


Figure 5.1: Meta-architecture of a two-layered UNet, GeoDUDe-2, with Transfer Branch used in deep learning algorithms. For 2D problems the input is upsampled along the receiver axis by deconvolutions in the Transfer Branch. UNet's have "rungs" that connects the encoder and decoder branches. In this way, the network can incorporate both low and high resolution data.

It is suggested in [36, 83] that such a bottleneck can reduce the number of necessary feature maps and so improve computational efficiency. We use Rectified Linear Units (ReLUs) [30] for our activation and size 3x3 (3x1 for 1D interface problems) filter kernels for our convolutions. As in [44], we use Batch-Normalization [46] to help smooth training. The setup is shown in Figure 5.2

Dense Blocks. Though GeoSeg supports multiple block types, all the networks reported in this paper use dense blocks. These are stacks of convolutional layers as shown in Figure 5.2. The defining features of these blocks, introduced in [44] is that every layer receives input from all previous layers in the block via concatenation. Such architectures have been shown to greatly improve results in image classification while reducing computational burden [44].

Transfer Branch. All of our meta-architectures preserve resolution of their input and so our detection resolution is limited by input resolution. This is not a problem in the temporal axis, which translates to the z axis in output, since we have a large number



(a) Convolutional Layer (b) Dense Block

Figure 5.2: The type of blocks used in GeoSeg for this paper: (a) Block compositions of a basic convolutional layer using a bottleneck convolution to expand the filter channels before the full convolution; (b) a corresponding dense block. Each layer of the block recieves input from all previous layers allowing information to flow through the whole block.

of time samples; however, the x-axis resolution is limited by the number of receivers we have for our input. To increase the resolution in this direction, we place a small l-layer CNN before the main network which upsamples the receiver axis, via strided deconvolutions, by a factor of 2^l.

5.2 Numerical Experiments

Here we present the performance of deep learning algorithms for the three detection experiments: 1D interface problme, three-layered media model, and a 2D single cylindrical pocket model. The architecture used for all experiments is a UNet with Dense Blocks (GeoDUDe). Each dense block will be made of four constituent bottle-necked convolutional layers with a bottle neck factor of 4. For all 1D networks the dense blocks'

convolutions use a kernel size of 3×1 in the base of the block and 2×1 at each transition layer, while for the 2D networks a 3×3 kernel size is used in the base block with a 2×2 kernel size in the transition layer. The meta-architectures had 16 filter channels except for the 1D interface model with P-wave data which only used 4. Our primary evaluation metric is accuracy which is the number of correctly predicted pixels over total pixels. For the 2D pocket model, we will also consider the Intersection Over Union metric which better captures segmentation performance.

In the 2D pocket model, a two-layer transfer branch was used. Each layer was a convolution, two-strided in the receiver direction with a kernel size of 3×3 with 4 filter channels. During training, these layers had a drop out probability of 0.2.

The initial P-Wave data is generated with source function

$$f_j^k(\mathbf{x}) = \cos(k(x_j - x_{0,j})) \exp\big(-2k|\mathbf{x} - \mathbf{x}_0|^2\big), \tag{5.2}$$

and the P,S-Wave initial data is generated from the Green's function in section 3.3. The same initial forcing function (4.4) is used for the P-,S- wave datasets. The data is generated on the cluster, *POD*, with the same set up as in section 4.2. As the initial condition is independent of the wavespeed, only one wave packet decomposition needs to be computed and saved for all data points to be generated. This saves a tremendous amount of time as only the ODE system needs to be solved for various wavespeeds and interface heights. For example to generate the P-Wave data, when 804672 total beams are used, each data point is generated in approximately 2.5 minutes. This is compared to SPECFEM3D which takes is approximately 45 minutes to generate a data point. All of the networks were trained on the Google Cloud Platform, or on the cluster *POD* with Keras 2.2.2 and Tensorflow 1.10.0 as a backend using a single NVIDIA Tesla V100 GPU.

5.2.1 1D Interface

To provide a proof of concept we first experimented with a two-layered flat interface model. We also use this case to investigate whether our network is simply inverting the FGA by comparing performance of a network trained on FGA but evaluated on data generated by SEM.

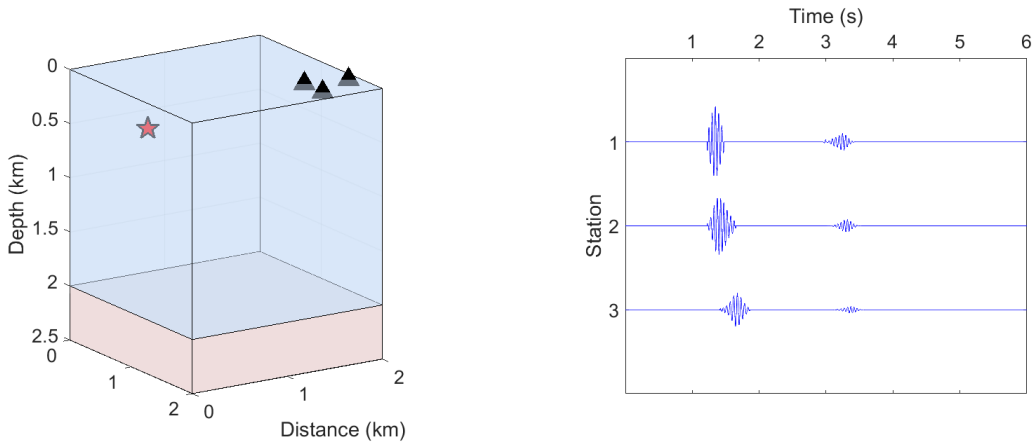
P-Wave Data

Dataset. The P-wave data set is generated with a computation domain of $[0, 2]$ km$\times[0, 2]$ km$\times[0, 2.5]$ km with a source centered at $\mathbf{x}_0 = (0.5, 0.5, 0.5)$ km and a nondimensionalized wavenumber $k = 1/\epsilon = 128$ in (5.2), which corresponds to approximately 20.37 Hz. The stations are located on the surface at $S_1 : (1.5, 1.5, 0)$ km, $S_2 : (1.8, 1.5, 0)$ km, $S_3 : (1.6, 1.9, 0)$ km. The interface is a plane, $z = z_0$ that varies from depth 1 km to 2.5 km. Above the interface the wavespeed varies from .78 km/s to 1.22 km/s, below the interface the wavespeed varies from 1.29 km/s to 1.56 km/s. See Figure 5.3.

Each data point is a $(6000, 3, 3)$ tensor. Prior to training, we further down sample the temporal dimension by a factor of 25 and normalize the amplitude of the seismogram data. There were a total of 7790 examples. The mini-batch size during training was 256 examples.

Network Details. As described above our architecture was a 1D GeoDUDe-3 where each convolutional layer in the dense block had 4 feature channels. The During training the dropout probability was set to 0.5 and a NADAM optimizer was used with default parameters.

Results. Network evaluations were performed with data generated by the FGA and



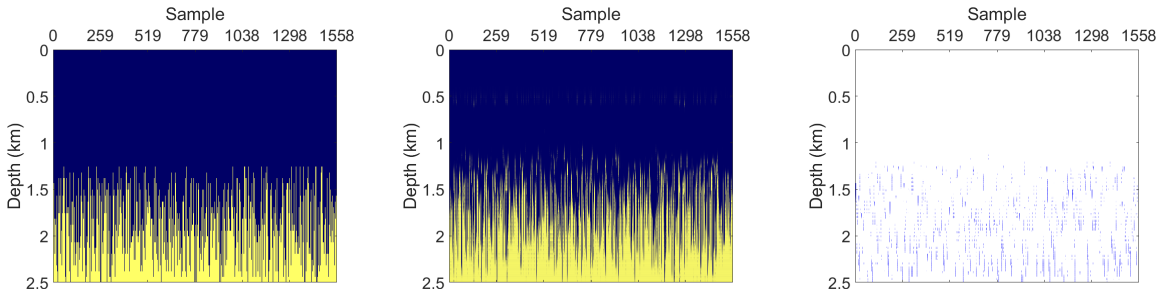
(a) Computational Domain

(b) Seismogram

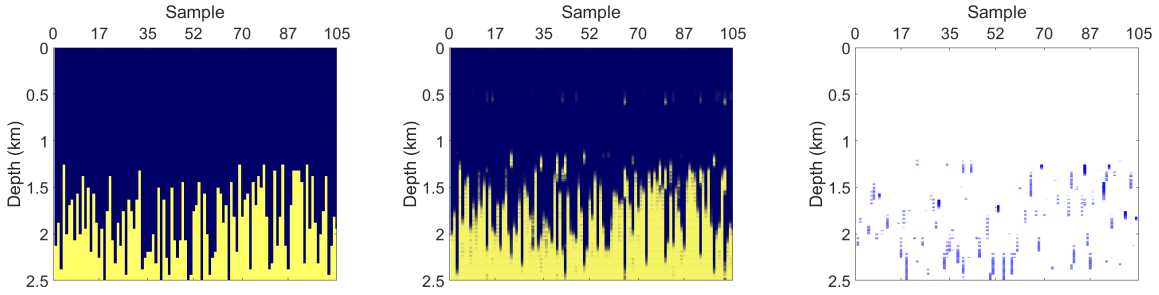
Figure 5.3: The locations of source and receivers, and the generated synthetic P-wave seismograms for the 1D interface problem. We take $k = 128$ for generating the synthetic data. (a) The source is located at (.5,.5,.5) km as a star and the 3 receivers are located on the surface. The interface presented is at a depth of 2 km. (b) A visualization of typical data point, which is a collection of 3 seismograms from the forward simulation using the FGA.

SPECFEM. Notably, the networks are never trained on any SPECFEM data. This was to investigate whether the network was sensitive to the asymptotic error produced by the FGA.

After 3500 epochs of training GeoDUDe-3 achieved a 96.97% evaluation accuracy on data generated by the FGA. When evaluated on data generated by SPECFEM dataset GeoDUDe-3 achieved a 94.29% evaluation accuracy, only a 2.68% decrease. We remark in [63, 86], it was shown even small perturbations in input can affect network classification results. This suggests that the asymptotic errors present in the FGA do not greatly affect the segmentation problem. Visualizations of the output for GeoDUDe-3 are shown in Figure 5.4. Figure 5.5 shows the heatmap. Recall this displays the confidence the network places on the pixels prediction.

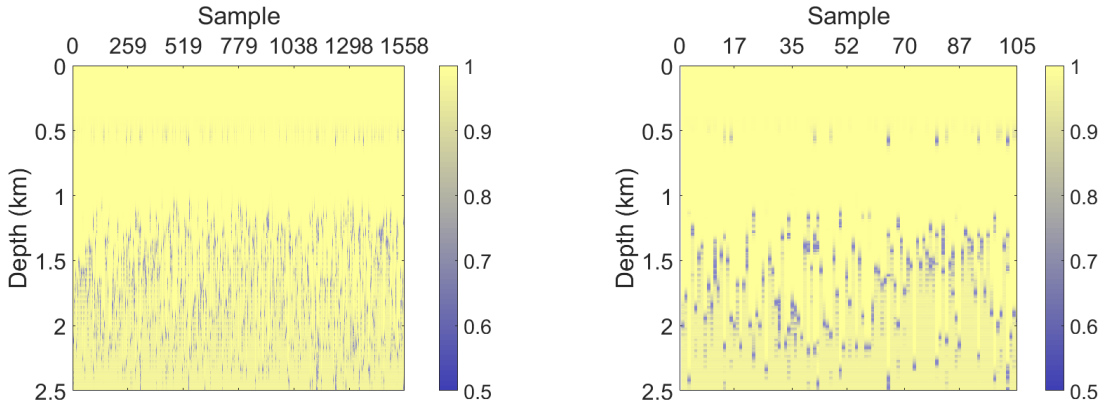


(a) FGA: Actual (b) FGA: Predicted (c) FGA: Difference



(d) SEM: Actual (e) SEM: Predicted (f) SEM: Difference

Figure 5.4: 1D interface predicted by GeoDUDe-3 using P-wave data. Each column of pixels represents a sample. The value of each pixel describes whether the material at the depth corresponding to that pixel's column belongs to either the high or low wavespeed region.



(a) FGA (b) SEM

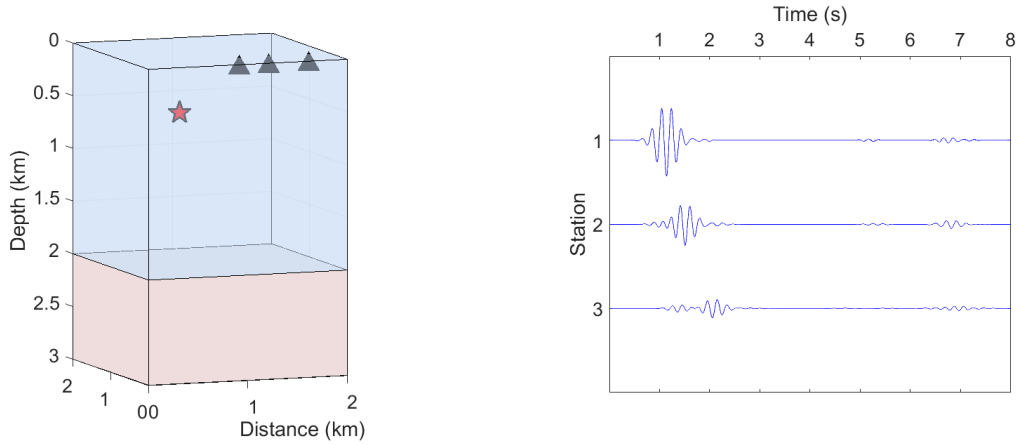
Figure 5.5: P-wave confidence distribution comparison produced by GeoDUDe-3 for 1D interface problem. Regions of low confidence correspond to areas where an interface is likely.

P,S-wave data set

Dataset. The P,S-wave dataset is generated with a computation domain of $[0, 2]$ km$\times[0, 2]$ km$\times[0, 3]$ km with a source centered at $\mathbf{x}_0 = (0.5, 0.5, 0.5)$ km, and wavenumber $k = 32$, or approximately 5.09 Hz. The stations lie in a plane and are located just below the surface at $S_1 : (1.1, 0.5, 0.1)$ km, $S_2 : (1.4, 0.5, .1)$ km, $S_3 : (1.8, 0.5, 0.1)$ km. The interface is a plane, $z = z_0$ that varies from depth 1 km to 2 km. Above the interface c_p varies from 0.75 km/s to 1.10 km/s, below the interface c_p varies from 1.12 km/s to 1.48 km/s and we fix $c_\mathrm{s} = c_\mathrm{p}/1.7$ (corresponding the case $\lambda \approx \mu$). See Figure 5.6. There are a total of 6,400 data points in the P,S-wave dataset. Each data point is a (2048,3,3) tensor. Prior to training each example is down-sampled along the temporal axis by a factor of 8. Each network used a mini-batch training size of 256. Similarly to the P-wave dataset, 100 additional samples were generated using SPECFEM3D for evaluation after training.

Network Details. GeoDUDe-2 and GeoDUDe-3 with default parameters were used. Both networks were trained using a NADAM optimizer with a dropout probability of 0.5.

Results. Both networks were trained for 3500 epochs. The most successful network was GeoDUDe-2, with 98.26 % evaluation accuracy on FGA data, and 97.55 % evaluation accuracy on the SPECFEM data . We find that the evaluation accuracy goes down for deeper networks. In particular, GeoDUDe-3 performed worse with only a 92.34 % evaluation accuracy, especially compared to the same network architecture on the P-wave dataset. This is likely due to overfitting of the data causing an increase in generalization error. Similarly to the P-wave dataset, evaluation accuracies on SPECFEM3D data are only marginally worse than their FGA counterparts, with a max difference of 1.17% between the datasets. See Table 5.1 for the summary of the results and Figures 5.7, 5.8 and 5.9.

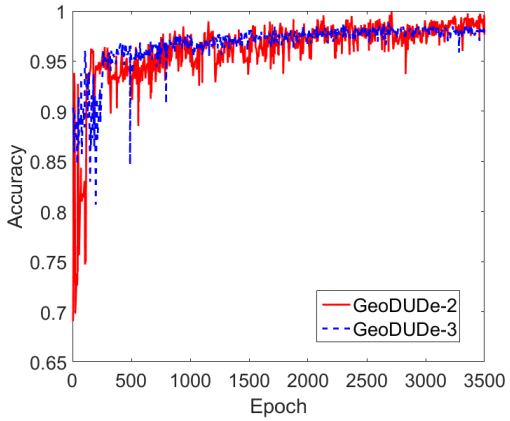


(a) Computational Domain (b) Seismogram

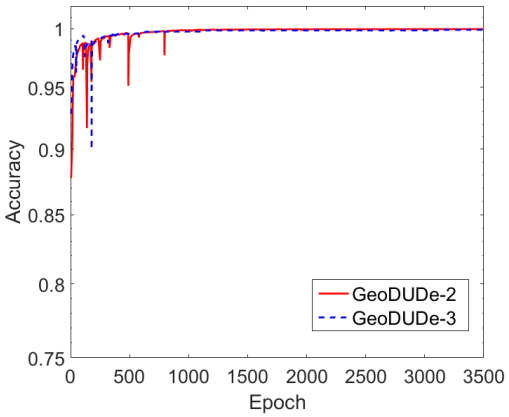
Figure 5.6: The locations of source and receivers, and the generated synthetic P- and S-wave seismograms for the 1D interface problem. We take $k = 32$ for generating the synthetic data. (a) The source is located at (.5,.5,.5) as a star and the 3 receivers are located on the surface. The interface presented is at a depth of 2 km. (b) A visualization of typical data point, which is a collection of 3 seismograms from the forward simulation using the FGA.

Network	**Eval. Acc.**	**Train Acc.**	**SEM Acc.**
GeoDUDe-2	98.26 %	99.97 %	97.55 %
GeoDUDe-3	97.64 %	99.90 %	96.47 %

Table 5.1: P,S-Data Network Comparisons for 1D interface problem. The columns represent evaluation accuracy, training accuracy, and evaluation accuracy tested by SEM synthetic data.

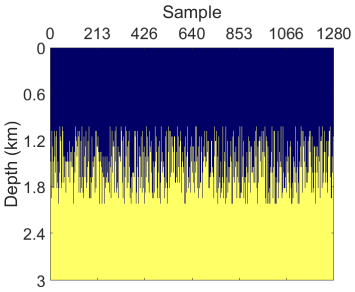


(a) Evaluation Accuracy

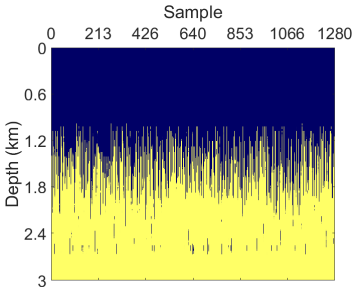


(b) Training Accuracy

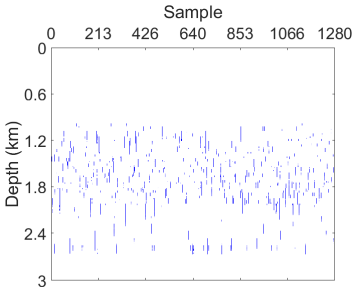
Figure 5.7: PS-wave training results for 1D interface problem, with synthetic data generated for $k = 32$ in (5.2): The evaluation data set for this figure only contains data generated by the FGA.



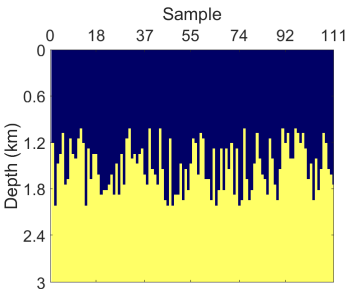
(a) FGA: Actual



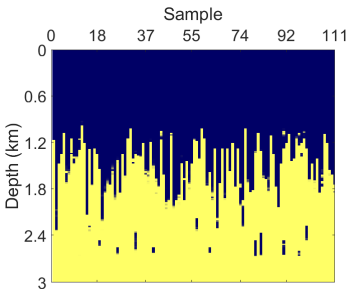
(b) FGA: Predicted



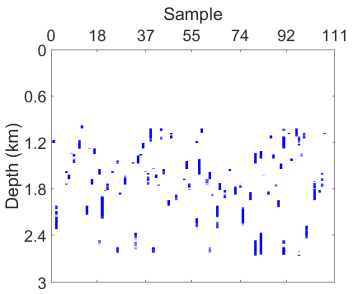
(c) FGA: Difference



(d) SEM: Actual

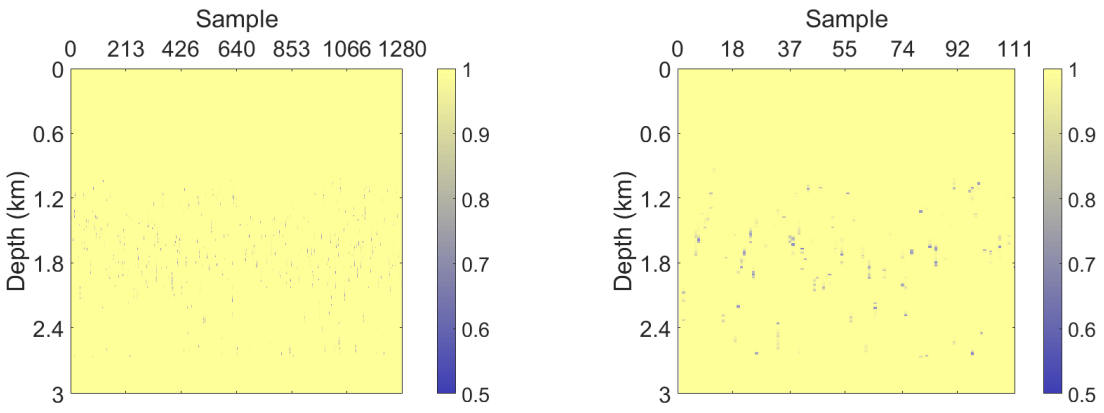


(e) SEM: Predicted



(f) SEM: Difference

Figure 5.8: 1D interface predicted by GeoDUDe-2 using P,S-wave data. Each column of pixels represents a sample. The value of each pixel describes whether the material at the depth corresponding to that pixel's column belongs to either the high or low wavespeed region.



(a) GeoDUDe-2:FGA

(b) GeoDUDe-2:SEM

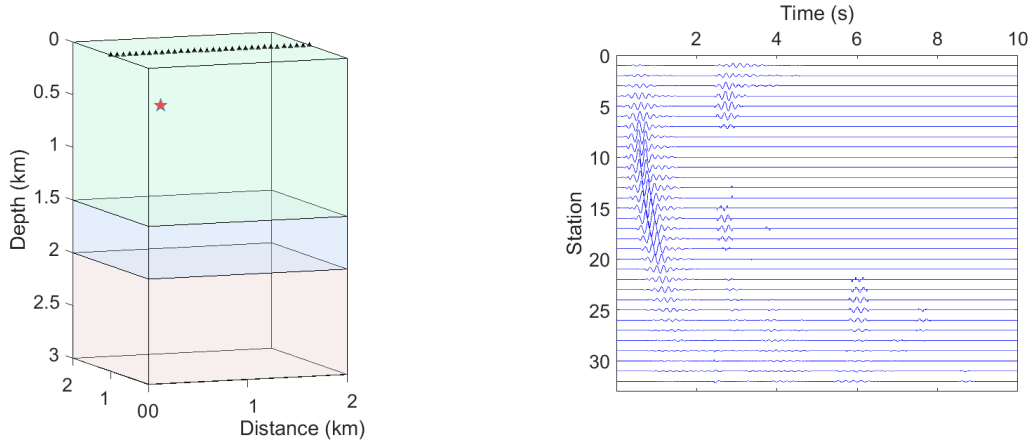
Figure 5.9: P,S-wave heat-map distribution comparison produced by GeoDUDe-2 for 1D interface problem. Regions of low confidence correspond to areas where an interface is likely.

Three-Layered Media

Dataset. A natural extension of the model is to include one or more low velocity regions in the computational domain. For this experiment we consider a three-layered media with a low velocity region in the middle, the velocities in each region will be fixed.The P-wave speed is $c_{\mathrm{p}} = 1.3, 0.9, 1.7$ km/s for the top, middle, and bottom layers, receptively. The S-wave speed is set to $c_{\mathrm{s}} = c_{\mathrm{p}}/1.7$ for each layer. The lower interface will be in a rage of 1.8 km and 2.8 km by an increment of 1 m. Similarly the upper interface will vary from .2 km to 1.2 km by an increment of 1m. See Figure 5.10. There were 10201 samples with a batch size of 64.

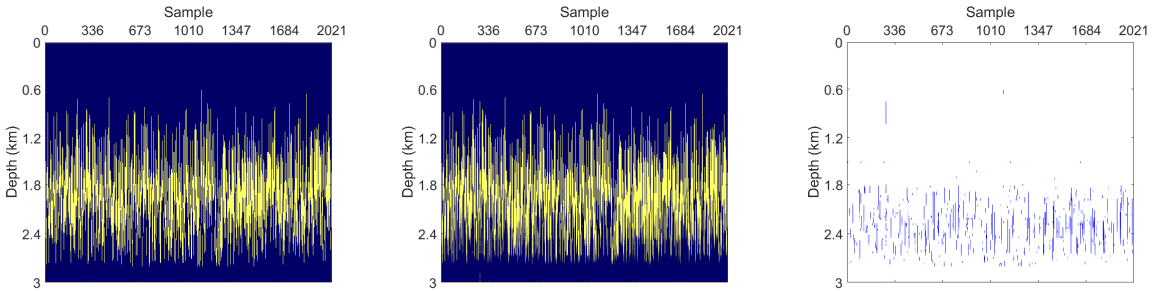
Network Details. GeoDUDe-3 was used. During training the dropout probability was 0.12. Training was performed with stochastic gradient descent with a learning rate of 0.001.

Results. The network achieved a training accuracy of 99.51% and an evaluation accuracy of 95.51% after 3000 epochs. See Figures 5.11 and 5.12.



(a) Computational domain (b) seismograph

Figure 5.10: The locations of source and receivers, and the generated synthetic P- and S-wave seismograms for the three-layered media model. We take $k = 32$ for generating the synthetic data. (a) The source is located at $(.5, 1, .5)$ km as a star, the 32 receivers are located on the surface on the plane $y = 1$ km, and the interfaces presented are at a depth of 1.5 km and 2 km. (b) A visualization of typical data point, which is a collection of 32 seismograms from the forward simulation using the FGA.



(a) Actual (b) Predicted (c) Difference

Figure 5.11: Predictions for three-layered media by GeoDUDe-3: Each column of pixels represents a sample. The value of each pixel describes whether the material at the depth corresponding to that pixel's column belongs to either the high or low wavespeed region. There is a slight loss of confidence for the network detecting the lower interface.

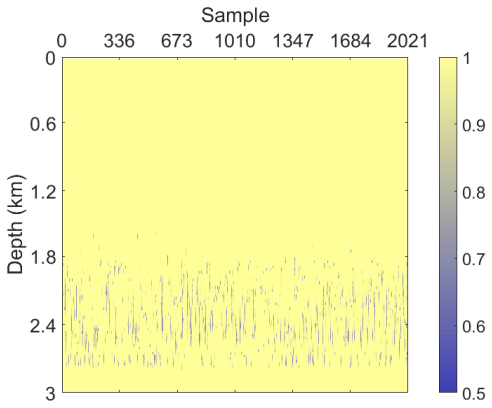
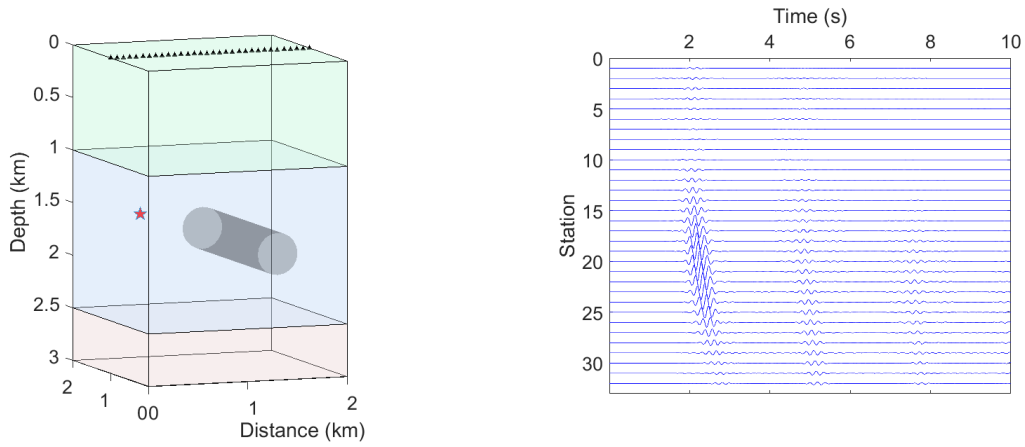


Figure 5.12: Confidence map for three-layered media model produced by GeoDUDe-3. Regions of low confidence correspond to areas where an interface is likely.

5.2.2 2D Low Velocity Pocket

Dataset: We now investigate whether the network can learn more complex 2D geometries. The considered models each will be a three-layered problem with a low velocity cylindrical region in the middle layer. The source will be located at $(.5, 1, 1.5)$ km. The interfaces located at 1 km and 2.5 km will be fixed. A cylinder with center (x, z) and radius r will be randomly generated $x \in [0.85, 1.65]$ km, $z \in [1.35, 2.15]$ km, and $r \in [.05, .3]$ km with samples taken from a uniform distribution. See Figure 5.13. 11350 data points are generated with 1000 being saved for evaluation. The P-wave speeds will be fixed and are $c_{\text{p}} = 1.1, 1.3, 1.7$ km/s, for the top, middle and bottom layers respectively. The S-wave speed, c_{s} will be a fixed multiple of c_{p} by 1.7 for each layer. Inside the pocket the P-wave speed is set to $c_{\text{p}} = 0.5$ km/s and the S-wave speed is set to zero, $c_{\text{s}} = 0$. Only P-waves will propagate through the cylinder; However, S-wave can transmit to P-wave going in the pocket and P-wave can transmit to P,S-waves coming out of the pocket. Unlike previous models the goal is to identify a low velocity region in a three layered media in a 2D slice of the computational domain. A batch size of 20 examples was used.

Network Design. A GeoDUDe-4 network was used with a two layer transfer branch



(a) Computational domain

(b) Seismogram

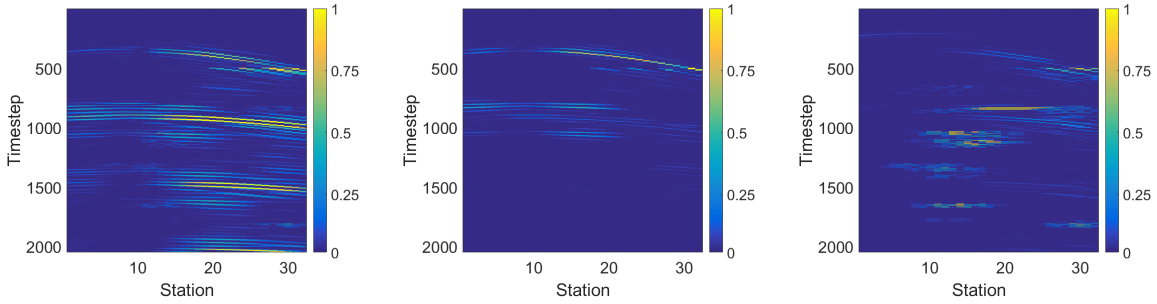
Figure 5.13: The locations of source and receivers, and generated synthetic P- and S-wave seismograms for the 2D pocket model. We take $k = 32$ for generating the synthetic data. (a) The source is located at $(.5, 1, 1.5)$ km as a star and the 3 receivers are located on the surface on the plane $y = 1$ km. The interfaces are fixed at a depth of 1 km and 2.5 km. (b) A visualization of typical data point, which is a collection of 32 seismograms from the forward simulation using the FGA.

before its input. The dropout probability was 0.2.

Results. The network achieved a training accuracy of 99.95% and an evaluation accuracy of 99.73% after 1428 epochs. In Figure 5.16 we see the networks are indeed learning geometry. This is particularly interesting given that the network only "sees" images like Figures 5.14 and 5.15. These results suggest the network is transforming the data in some way which we hope to explore in future work.

5.2.3 Effect of Noisy data

We now consider the 2D pocket example with additive white noise. Normally, noise is added to the training data set to increase the size of the set and lead to a more robust network. We take an evaluation set of 1000 data points and add *i.i.d.* (independent identically distributed) Gaussian noise to each time step of the displacement field data.



(a) x-coordinate (b) y-coordinate (c) z-coordinate

Figure 5.14: Visualization of network input using normalized displacement data for 2D pocket model.

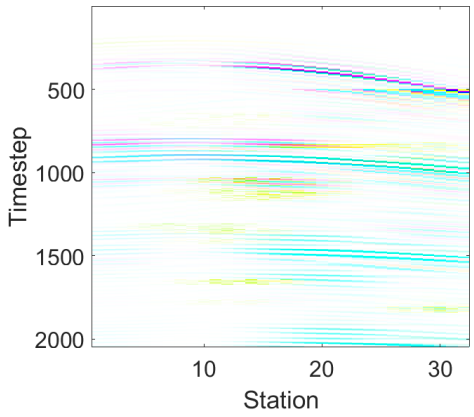
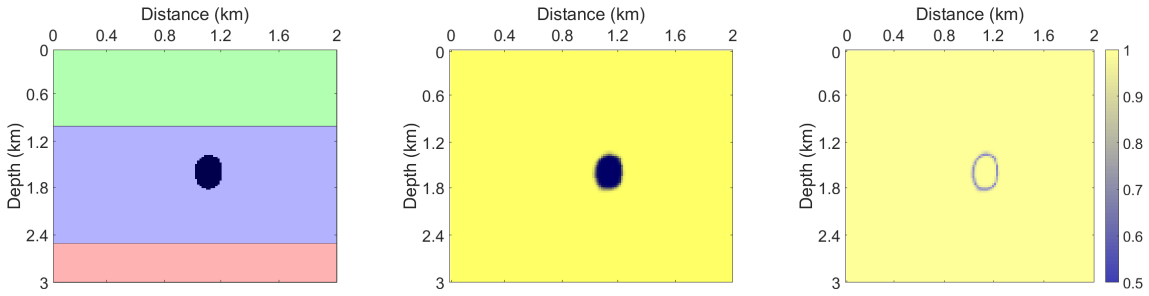
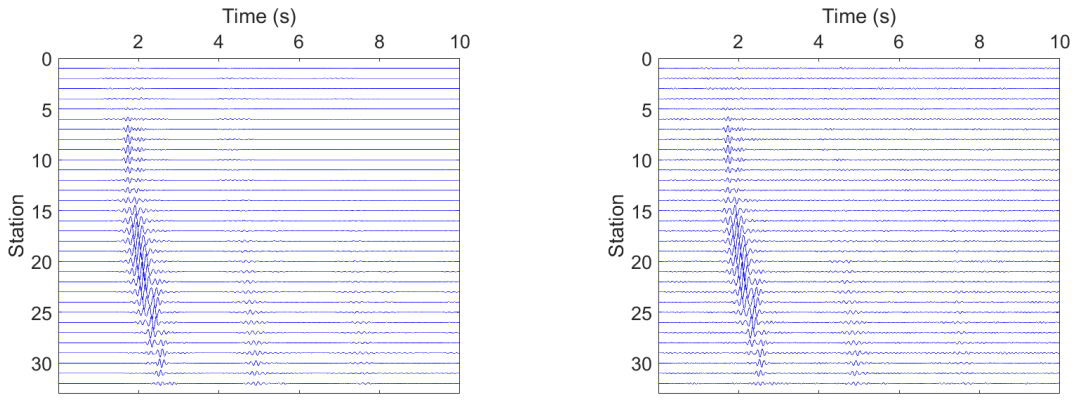


Figure 5.15: Visualization of network input as image for 2D pocket model. Each color channel (inverse RGB) represents a coordinate of the displacement.



(a) Actual (b) Predicted (c) Confidence map

Figure 5.16: 2D pocket results predicted by GeoDUDe-4, with a typical data point chosen for visualization. The pocket is recovered with the networks confidence wavering on the boundary of the pocket.



(a) No noise

(b) Noise strength of 1% of maximum recorded data

Figure 5.17: Comparison of seismograms with noise and no noise for the 2D pocket model. (a) Seismogram with no noise. (b) Additive Gaussian white noise at 1% of $\max|\mathbf{u}|$. This shows that 1% of the maximum recorded displacement is enough to mask the reflected data from the pocket.

For an individual data point, the noise strength can be calculated by

$$W_i = \frac{\sigma}{R \max|\mathbf{u}_r|}, \tag{5.3}$$

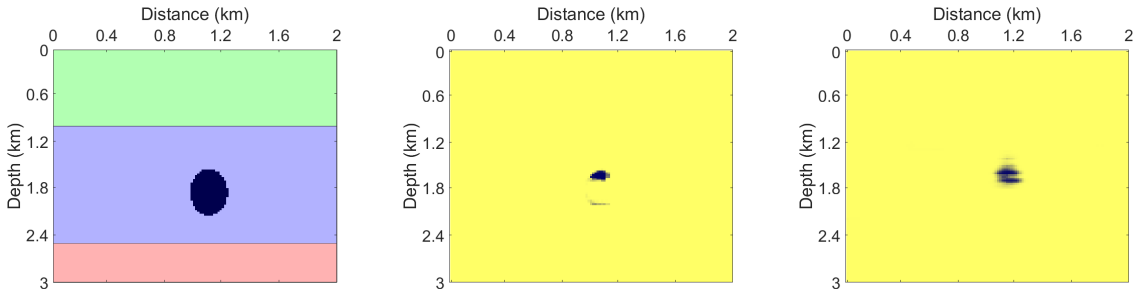
where R is the reflection coefficient and $\max|\mathbf{u}_r|$ is the maximum displacement from the reflected wave. The noise strength will be given by W, which is the approximate average value of W_i across the data set. The standard deviation σ is chosen so that W can be interrupted as a percentage of the reflected wave displacement, e.g., $W = 20$ gives of the a noise strength of 20% of the average max displacement of the reflected wave. We notice that with noise generated with a strength of 1% of the maximum of direct recorded displacement, the reflected data from the pocket is the same order of magnitude of the noise, effectively masking it. See Figure 5.17.

Network Design: To compare results, we use the same model as in the previous Section 5.2.2 and train a network with the same parameters, with a noise strength of $W = 20$.

	Unperturbed	$W = 10$	$W = 50$
Trained without Noise	0.8163	0.7335	0.1308
Trained with Noise	0.8706	0.7576	0.5249

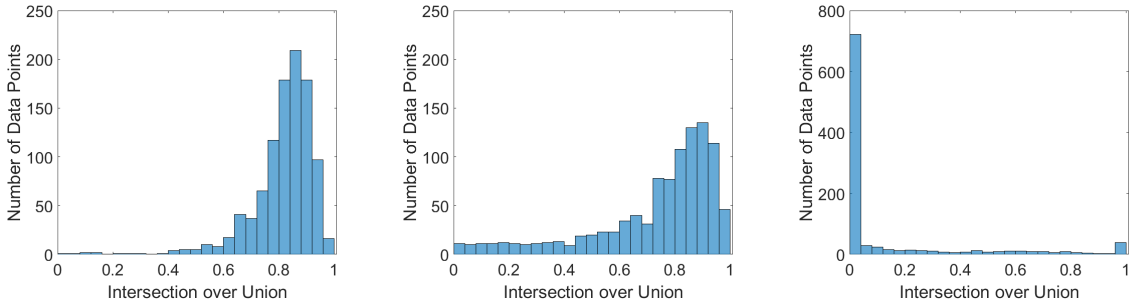
Table 5.2: IOU Scores for GeoDUDe-4 trained with and without noise for the 2D pocket model.

Results. A GeoDUDe-4 was trained for 2000 epochs with additional noise for a final evaluation accuracy of 99.731% evaluation accuracy. However, evaluation accuracy can be a misleading metric for network performance in pocket detection since assigning the high velocity class to every pixel could get an accuracy up to 80% on some samples. Instead intersection over union (IOU) is used (see [75] for a more detailed explanation). Figures 5.18 and 5.19 show the histograms the IOU scores of networks trained with and without white noise evaluated on the evaluation data with no additional noise, additional noise strength $W = 10$, and additional noise strength $W = 50$ respectively. While both networks display good IOU scores on the unperturbed data and when the data is only perturbed with noise strength $W = 10$, the benefits of additional noise in training become clear when the noise strength is increased to $W = 50$: the IOU scores of the network trained without noise on noisy data plummets, effectively misclassifying almost every pocket, while the IOU score of the network trained with noise decreases, but maintains many correct classifications. The average IOU scores are summarized in Table 5.2. Evaluating on higher noise strength collapses the network's output to no pocket detected.



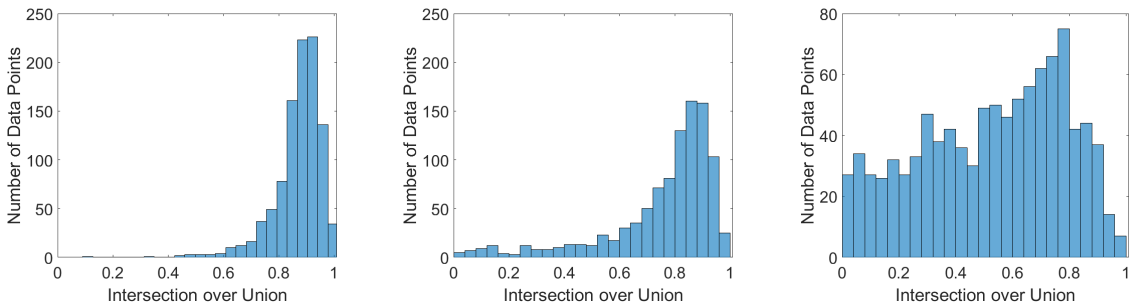
(a) ground truth (b) IOU=0.1403 (c) IOU=0.2052

Figure 5.18: Visualization of IOUs by GeoDUDe-4 for the 2D pocket model. Results taken from network trained with noise. Data is augmented with noise with a noise strength of 50%. (a) ground truth for comparision. (b) IOU=0.1403. (c) IOU= 0.2052. For each displayed results, the network was able to detect the location of the pocket. With additional noise the network is unable to resolve the geometry.



(a) No noise (b) W=10 (c) W=50

Figure 5.19: Network with trained without noise, 1000 data points are plotted in each Histogram. Subfigures (a), (b), (c) show the IOU metric with no noise, 10% noise strength, and 50% noise strength respectively.



(a) No noise (b) W=10 (c) W=50

Figure 5.20: Performance of the Network GeoDUDe-4 trained with noise strength at 20% of the average max displacement of the reflected wave for the 2D pocket model. 1000 data points are plotted in each Histogram. Subfigures (a), (b), (c) show the IOU metric, with no noise, 20% noise strength, and 50% noise strength, respectively.

Chapter 6

Concluding remarks

In this work, the FGA for high frequency elastic waves has been derived, proven and systematically verified to an effective modeling tool for seismology. The derivation uses an orthonormal frame based on the momentum variable and Geophysical properties. The initial conditions, transmission and reflection conditions for sharp interfaces (e.g., Moho surface and Core-mantle boundary), have been numerically verified to have the $\mathcal{O}(\epsilon)$ accuracy. A comparison with the FGA to SPECFEM3D with the forward modeling and has been shown to be accurate and a the developed API GeoSeg, shown minor differences between the data sets.

In the original mathematical work on FGA for strictly linear hyperbolic systems [59], the FGA formulation uses an eigenfunction decomposition. It is shown that the this formulation is asymptotically equivalent to the eigenfunction decomposition. The tools used, generalize the results in [59] as the elastic wave equation is not strictly hyperbolic. Other vector wave equations can be investigated with the methods used. By these efforts, it is hoped a bridge in the gap of semiclassical approximation theory in mathematics and the application of 3-D seismic tomography in geophysics can be realized. Thus, providing those working the field of Geophysics a tool for fast efficient modeling of seismic waves.

The use of the FGA to generate large amounts of seismic data provides a quick way to generate labeled synthetic data for statistical learning of the inverse tomography problem. Casting the inverse problem as a segmentation problem resulted in high evaluation accuracy networks for piecewise constant two-layer models on both FGA and SEM datasets. The UNet architectures with dense blocks displayed superior accuracy compared to simpler network architectures, however, deeper networks did not necessarily outperform their shorter counterparts. On the two layer benchmark problem the networks exhibited good invariance of prediction in regard to which numerical method was used to generate the dataset, likely because the FGA and SEM exhibit the same traveltime information. Having a network independent of numerical method is important, and the FGA can help to train such a network as it generates synthetic seismic data that carries the correct traveltime information of the real-world data. Further, analogous meta-architectures also exhibit high evaluation and IOU accuracy for pocket detection in noisy data.

6.1 Future Research

An immediate interest is to apply 3-D seismic tomography with the FGA to real data around their dominant frequencies. Also, numerical instabilities from section 4.2.3 may also be resolved by other strategies [62], and a further study on optimal transport theory-based seismic tomography might be fruitful. It will be interesting to study the performance of the FGA method in the optimal transport theory-based seismic tomography, with the normalization strategies proposed in, [22, 73, 102, 16] as well. Another interest is to apply this projection method to other high frequency vector wave equation; e.g. Maxwell's equations. Another idea that will provide useful is to cast an approximation for attenuation in the FGA framework. As integration by parts is required for simplification; nonlinearities are difficult but not all are impossible to deal with.

The success of the study with FGA datasets involving neural networks on substructure geometries act as a stepping stone to tackle more complicated and realistic geological models. By developing the API GeoSeg, it is easy to implement neural networks designed for the reported example models and more general segmentation problems of seismogram data than those discussed throughout this work. Together with the FGA, the task of training a deep neural network on sufficiently large amounts of seismogram data becomes a computationally affordable task. Immediate future directions to be explored are multi-pocket models, multi-nonlinear interface models with and without pockets present. Long term goal is to develop a neural network model to tackle fully 3D substructure geometries and develop a neural network trained on synthetic seismic data capable of making inferences from real seismic data.

Bibliography

[1] K. Aki, A. Christofferson, and E. Husebye. Determination of the three-dimensional seismic structure of the lithosphere. *J. Geophys. Res.*, 82:277–296, 1977.

[2] K. Aki and W. Lee. Determination of the three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes 1. A homogeneous intial model. *J. Geophys. Res.*, 81:4381–4399, 1976.

[3] K. Aki and P. Richards. *Quantitative Seismology*. Geology Seismology. University Science Books, 2002.

[4] M. Araya-Polo, T. Dahlke, C. Frogner, C. Zhang, T. Poggio, and D. Hohl. Automated fault detection without seismic processing. *The Leading Edge*, 36(3):208–214, 2017.

[5] M. Araya-Polo, J. Jennings, A. Adler, and T. Dahlke. Deep-learning tomography. *The Leading Edge*, 37(1):58–66, 2018.

[6] A. Bach. *An Introduction to Semiclassical and Microlocal Analysis*. Universitext. Springer New York, 2002.

[7] T. Belytschko and R. Mullen. On dispersive properties of finite element solutions. In K. J. Marfurt, editor, *Numerical modeling of seismic wave propagation*, pages 336–351. Soc. of Expl. Geophys., 13 edition, 1990. Reprinted from Modern problems in elastic wave propagation, Miklowitz and Achenback, Eds., 67-82.

[8] M. Berry. Quantal phase factors accompanying adiabatic changes. *Proc. Roy. Soc. Lond. A*, 392:45–57, 1984.

[9] A. Bourgeois, M. Bourget, P. Lailly, M. Poulet, P. Ricarte, R. Versteeg, and G. Grau. Marmousi, model and data. *The Marmousi Experience*, pages 5–16, 1991.

[10] V. Cerveny, M. M. Popov, and I. Psencik. Computation of wave fields in inhomogeneous media – Gaussian beam approach. *Geophys. J. Roy. Astr. Soc.*, 70:109–128, 1982.

[11] L. Chai, P. Tong, and X. Yang. Frozen Gaussian approximation for 3-D seismic wave propagation. *Geophysical Journal International*, 208(1):59–74, 2017.

[12] L. Chai, P. Tong, and X. Yang. Frozen Gaussian approximation for 3-D seismic tomography. *Inverse Problems*, 34:055004, 2018.

[13] C. Chapman. Exact and approximate ray theory in vertically inhomogeneous media. *Geophys. J. R. astr. Soc.*, 6:201–233, 1976.

[14] C. Chapman and R. Drummond. Body-wave seismograms in inhomogeneous media using Maslov asymptotic theory. *Bull. Seism. Soc. Am.*, 72:277–317, 1982.

[15] Y. V. Chekanov. Caustics in geometrical optics. *Functional Analysis and Its Applications*, 20(3):223–226, Jul 1986.

[16] J. Chen, Y. Chen, H. Wu, and D. Yang. The quadratic Wasserstein metric for earthquake location, 2017. arXiv:1710.10447.

[17] Y. H. Chu, J. I. Huang, and V. Sonnad. 3-D forward modeling by the pseudospectral method on a coupled array of processors. In *58th Ann. Internat. Mtg*, page Session:S23.6. Soc. of Expl. Geophys., 1988.

[18] M. V. de Hoop and A. T. de Hoop. Spectral-domain parabolic approximation and its consequence for the space-time scalar wave emitted by a point source. In *60th Ann. Internat. Mtg*, pages 1045–1048. Soc. of Expl. Geophys., 1990.

[19] R. Delgadillo, J. Lu, and X. Yang. Gauge-invariant frozen gaussian approximation method for the schrödinger equation with periodic potentials. *SIAM Journal on Scientific Computing*, 38(4):A2440–A2463, 2016.

[20] R. A. Delgadillo. Semiclassical methods for high frequency wave propagation in periodic media. Jan 2016.

[21] T. Dozat. Incorporating nesterov momentum into adam. 2016.

[22] B. Engquist and B. Froese. Application of the wasserstein metric to seismic signals. *Commun. Math. Sci.*, 12:979–988, 2014.

[23] B. Engquist and O. Runborg. Computational high frequency wave propagation. *Acta Numer.*, 12:181–266, 2003.

[24] A. Fichtner. *Full Seismic Waveform Modelling and Inversion*. Springer Verlag, 2010.

[25] G. B. Folland. *Harmonic analysis in phase space*. Annals of Mathematics Studies, no. 122. Princeton University Press, Princeton, 1989.

[26] S. Fomel and N. Tanushev. Time-domain seismic imaging using beams. In *79th Ann. Internat. Mtg*, pages 2747–2752. Soc. of Expl. Geophys., 2009.

[27] B. Fornberg. The pseudospectral method: Comparisons with finite differences for the elastic wave equation. In *56th Ann. Internat. Mtg.*, page Session:S17.5. Soc. of Expl. Geophys., 1986.

[28] R. Gans. Fortpflanzung des lichts durch ein inhomogenes medium. *Annalen der Physik*, 352(14):709–736, 1915.

[29] R. Geller, H. Mizutani, N. Takeuchi, and N. Hirabayashi. Optimally Accurate Finite Difference Operators and Their Application to Forward and Inverse Seismic Modeling. In *64th Mtg.*, page E013. Eur. Assn. Geosci. Eng., 2002.

[30] X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks. In *AISTATS*, 2011.

[31] S. Gray and N. Bleistein. True-amplitude gaussian-beam migration. *Geophysics*, 74(2):S11–S23, 2009.

[32] S. H. Gray. Efficient traveltime calculations for Kirchhoff migration. *Geophysics*, 51:1685–1688, 1986.

[33] S. H. Gray. Gaussian beam migration of common-shot records. *Geophysics*, 70:S71–S77, 2005.

[34] J. C. Hateley, J. Roberts, K. Mylonakis, and X. Yang. Deep Learning Seismic Substructure Detection using the Frozen Gaussian Approximation. *arXiv e-prints*, page arXiv:1810.06610, Oct 2018.

[35] J. C. Hateley, X. Yang, L. Chai, and P. Tong. Frozen Gaussian approximation for 3-D elastic wave equation and seismic tomography. *Geophysical Journal International*, 216(2):1394–1412, 11 2019.

[36] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[37] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[38] E. J. Heller. Frozen Gaussians: A very simple semiclassical approximation. *J. Chem. Phys.*, 75:2923–2931, 1981.

[39] D. Helmberger. The crust-mantle transition in the Bering sea. *Bull. Seism. Soc. Am.*, 58:179–214, 1968.

[40] D. Helmberger and S. Ni. Approximate 3D Body-Wave Synthetics for Tomographic Models. *Bull. Seism. Soc. Am.*, 95:212–224, 2005.

[41] M. F. Herman and E. Kluk. A semiclasical justification for the use of non-spreading wavepackets in dynamics calculations. *Chemical Physics*, 91(1):27 – 34, 1984.

[42] N. R. Hill. Gaussian beam migration. *Geophysics*, 55:1416–1428, 1990.

[43] N. R. Hill. Prestack Gaussian-beam depth migration. *Geophysics*, 66:1240–1250, 2001.

[44] G. Huang, Z. Liu, and K. Q. Weinberger. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016.

[45] D. Iagolnitzer. Appendix microlocal essential support of a distribution and decomposition theorems — an introduction. In F. Pham, editor, *Hyperfunctions and Theoretical Physics*, pages 121–132, Berlin, Heidelberg, 1975. Springer Berlin Heidelberg.

[46] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.

[47] IRIS DMC. Data Services Products: iasp91, P & S velocity reference Earth model, 2010.

[48] S. Ji, S. Sun, Q. Wang, and D. Marcotte. Lamé parameters of common rocks in the earth's crust and upper mantle. *Journal of Geophysical Research: Solid Earth*, 115(B6), 2010.

[49] K. Kay. Integral expressions for the semi-classical time-dependent propagator. *J. Chem. Phys.*, 100:4377–4392, 1994.

[50] T. H. Keho and W. B. Beydoun. Paraxial ray Kirchhoff migration. *Geophysics*, 53:1540–1546, 1988.

[51] B. Kennett. (complier and editor). IASPEI 1991 seismological tables, 1991. Bibliotech, Canberra, Australia.

[52] S. Kim. On accuracy of finite difference amplitudes and interpolated traveltimes. In *71st Ann. Internat. Mtg*, pages 1175–1178. Soc. of Expl. Geophys., 2001.

[53] D. Komatitsch and J. Tromp. Spectral-element simulations of global seismic wave propagation-i. validation. *Geophysical Journal International*, 149(2):390–412, 2002.

[54] Y. LeCun, Y. Bengio, and G. E. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[55] Y. Li, F. D. Doherty, and J. C. Jackson. VSp Velocity Analysis Using a Phase Spectral Method. In *65th Mtg.*, page F41. Eur. Assn. Geosci. Eng., 2003.

[56] Q. Liu and Y. J. Gu. Seismic imaging: From classical to adjoint tomography. *Tectonophysics*, 566-567:31–66, 2012.

[57] D. Loewenthal, C. J. Wang, O. G. Johnson, and C. Juhlin. High order finite difference modelling and reverse time migration. *Expl. Geophys.*, 22(03):533–545, 1991.

[58] J. Lu and X. Yang. Frozen Gaussian approximation for high frequency wave propagation. *Commun. Math. Sci.*, 9:663–683, 2011.

[59] J. Lu and X. Yang. Convergence of frozen Gaussian approximation for high frequency wave propagation. *Comm. Pure Appl. Math.*, 65:759–789, 2012.

[60] J. Lu and X. Yang. Frozen Gaussian approximation for general linear strictly hyperbolic systems: Formulation and Eulerian methods. *Multiscale Model. Simul.*, 10:451–472, 2012.

[61] Y. Luo and G. T. Schuster. Wave-equation traveltime inversion. *Geophysics*, 56(5):645–653, 05 1991.

[62] L. Métivier, R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux. An optimal transport approach for seismic tomography: application to 3D full waveform inversion. *Inverse Problems*, 32:115008, 2016.

[63] S. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard. Universal adversarial perturbations. *CoRR*, abs/1610.08401, 2016.

[64] M. Motamed and O. Runborg. Taylor expansion and discretization errors in Gaussian beam superposition. *Wave Motion*, 47:421–439, 2010.

[65] H. Nakamichi, H. Hamaguchi, S. Tanaka, S. Ueki, T. Nishimura, and A. Hasegawa. Source mechanisms of deep and intermediate-depth low-frequency earthquakes beneath iwate volcano, northeastern japan. *Geophysical Journal International*, 154(3):811–828, 2003.

[66] S. Ni, X. Ding, and D. Helmberger. Constructing synthetics from deep earth tomographic models. *Geophys. J. Int.*, 140:71–82, 2000.

[67] R. Nowack, M. Sen, and P. Stoffa. Gaussian beam migration for sparse common-shot and common-receiver data. In *73rd Ann. Internat. Mtg.*, pages 1114–1117. Soc. of Expl. Geophys., 2003.

[68] T. Perol, M. Gharbi, and M. Denolle. Convolutional neural network for earthquake detection and location. *Science Advances*, 4(2), 2018.

[69] M. M. Popov, N. M. Semtchenok, P. M. Popov, and A. R. Verdel. Depth migration by the Gaussian beam summation method. *Geophysics*, 75:S81–S93, 2010.

[70] R. Pratt and N. Goulty. Combining wave-equation imaging with travel-time tomography to from high-reso1ution images from cross-hole data. *Geophysics*, 56:208–224, 02 1991.

[71] R. G. Pratt and R. M. Shipp. Seismic waveform inversion in the frequency domain, part 2: fault delineation in sediments using crosshole data. *Geophysics*, 64:902–914, 1999.

[72] J. Qian and L. Ying. Fast Gaussian wavepacket transforms and Gaussian beams for the Schrödinger equation. *J. Comput. Phys.*, 229:7848–7873, 2010.

[73] L. Qiu, J. Ramos-Martínez, A. Valenciano, Y. Yang, and B. Engquist. Full-waveform inversion with an exponentially encoded optimal-transport norm. In *SEG Technical Program Expanded Abstracts 2017*, pages 1286–1290. Society of Exploration Geophysicists, 2017.

[74] A. Quintus-Bosz and G. T. Schuster. Wavepath eikonal traveltime inversion; theory. *Geophysics*, 58(9):1314–1323, 09 1993.

[75] M. A. Rahman and Y. Wang. Optimizing intersection-over-union in deep neural networks for image segmentation. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, F. Porikli, S. Skaff, A. Entezari, J. Min, D. Iwai, A. Sadagic, C. Scheidegger, and T. Isenberg, editors, *Advances in Visual Computing*, pages 234–244, Cham, 2016. Springer International Publishing.

[76] N. Rawlinson, S. Pozgay, and S. Fishwick. Seismic tomography: A window into deep Earth. *Phys. Earth Planet. Inter.*, 178(3-4):101–135, 2010.

[77] B. Romanowicz. Seismic tomography of the Earth's mantle. *Annu. Rev. Earth Planet. Sci.*, 19:77–99, 1991.

[78] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[79] O. Runborg. Mathematical models and numerical methods for high frequency waves. *Commun. Comput. Phys.*, 2:827–880, 2007.

[80] Y. Sasaki. Three-dimensional resistivity inversion using the finite-element method. In *62nd Ann. Internat. Mtg*, pages 423–426. Soc. of Expl. Geophys., 1992.

[81] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015.

[82] B. Simon. Holonomy, the quantum adiabatic theorem, and Berry's phase. *Phys. Rev. Lett.*, 51:2167, 1983.

[83] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. A. Riedmiller. Striving for simplicity: The all convolutional net. *CoRR*, abs/1412.6806, 2014.

[84] I. Stekl and R. G. Pratt. Accurate seismic visco-elastic frequency-domain finite differences using rotated operators. In *58th Mtg.*, page Session:C041. Eur. Assn. Geosci. Eng., 1996.

[85] T. Swart and V. Rousse. A mathematical justification of the Herman-Kluk propagator. *Commun. Math. Phys.*, 286:725–750, 2009.

[86] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. J. Goodfellow, and R. Fergus. Intriguing properties of neural networks. *CoRR*, abs/1312.6199, 2013.

[87] N. M. Tanushev, R. Tsai, S. Fomel, and B. Engquist. Gaussian beam decomposition for seismic migration, 2011. ICES Report: 11-08.

[88] E. Tessmer and D. Kosloff. 3-D elastic modeling with surface topography by a Chebyshev spectral method. In *61st Ann. Internat. Mtg*, pages 1565–1568. Soc. of Expl. Geophys., 1991.

[89] E. Tessmer, G. Tessmer, D. Kosloff, and A. Behle. 3-D elastic modeling by a Chebychev spectral method. In *60th Ann. Internat. Mtg*, pages 1049–1052. Soc. of Expl. Geophys., 1990.

[90] P. Tong, D. Yang, D. Li, and Q. Liu. Time-evolving seismic tomography: The method and its application to the 1989 loma prieta and 2014 south napa earthquake area, california. *Geophysical Research Letters*, 44(7):3165–3175, 2017.

[91] J. Tromp, C. Tape, and Q. Liu. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophys. J. Int.*, 160(1):195–216, 2005.

[92] A. Vafidis and E. R. Kanasewich. Modelling crosshole seismic data in steam injection problems with finite differences. *J. Can. Soc. Expl. Geophys.*, 27(01):23–33, 1991.

[93] J. Vidale and D. Helmberger. Elastic finite-difference modeling of the 1971 San-Fernando, California earthquake. *Bull. Seism. Soc. Am.*, 78:122–141, 1988.

[94] J. Virieux. SH-wave propagation in heterogeneous media - Velocity-stress finite-difference method. *Geophysics*, 49(11):1933–1942, 1984.

[95] J. Virieux. P-Sv wave propagation in heterogeneous media: Velocity-stress finite-difference method. In K. J. Marfurt, editor, *Numerical modeling of seismic wave propagation*, pages 177–189. Soc. of Expl. Geophys., 13 edition, 1990. Reprinted from Geophysics, 51, 889-901.

[96] J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74:WCC1–WCC26, 2009.

[97] D. Wei and X. Yang. Eulerian gaussian beam method for high frequency wave propagation in heterogeneous media with discontinuities in one direction. *Commun. Math. Sci.*, 10:1287–1299, 2012.

[98] D. Witte and P. G. Richards. Contributions to the pseudospectral method for computing synthetic seismograms. In *57th Ann. Internat. Mtg*, page Session:S6.1. Soc. of Expl. Geophys., 1987.

[99] Y. Wu, Y. Lin, Z. Zhou, D. C. Bolton, J. Liu, and P. Johnson. Deepdetect: A cascaded region-based densely connected network for seismic event detection. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–14, 2018.

[100] G. Xie, K. H. Lee, J. Li, L. Pellerin, and D. Zuo. 3-D fast finite element born accelerating electromagnetic imaging using integral equation. In *66th Ann. Internat. Mtg*, pages 261–264. Soc. of Expl. Geophys., 1996.

[101] X. Yang, J. Lu, and S. Fomel. Seismic modeling using the frozen Gaussian approximation. *SEG Technical Program Expanded Abstracts 2013*, pages 4677–4682, 2013.

[102] Y. Yang, B. Engquist, J. Sun, and B. Hamfeldt. Application of optimal transport and the quadratic wasserstein metric to full-waveform inversion. *Geophysics*, 83:R43–R62, 2018.

[103] O. Yilmaz. *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*. Society of Exploration Geophysicists, 2001.

[104] K. Žáček. *Smoothing the Marmousi Model*, pages 1507–1526. Birkhäuser Basel, Basel, 2002.

[105] Z. Zhang, Q. Liu, and Y. Wang. Road extraction by deep residual u-net. *CoRR*, abs/1711.10684, 2017.

[106] D. Zhao. Tomography and dynamics of Western-Pacific subduction zones. *Monogr. Environ. Earth Planets*, 1:1–70, 2012.

[107] W. Zhu and G. C. Beroza. PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1):261–273, 10 2018.

[108] X. Zhu and G. A. McMechan. Estimation of a two-dimensional seismic compressional-wave velocity distribution by iterative tomographic imaging. *International Journal of Imaging Systems and Technology*, 1(1):13–17, 1989.

[109] R. W. Zimmerman, W. H. Somerton, and M. S. King. Compressibility of porous rocks. *Journal of Geophysical Research: Solid Earth*, 91(B12):12765–12777, 1986.